DOCUMENT RESUME

ED 457 306 UD 034 458

TITLE	The Longitudinal Evaluation of School Change and Performance (LESCP) in Title I Schools. Final Report. Volume 2: Technical Report.
INSTITUTION	Westat, Inc., Rockville, MD.; Policy Studies Associates, Inc., Washington, DC.
SPONS AGENCY	Department of Education, Washington, DC. Planning and Evaluation Service.
REPORT NO	PES-2001-20
PUB DATE	2001-00-00
NOTE	136p.; For Volume 1 (Executive Summary), see UD 034 457.
CONTRACT	EA-96008001
AVAILABLE FROM	ED Pubs, P.O. Box 1398, Jessup, MD 20794-1398. Tel: 877-433-7827 (Toll Free); Tel: 800-872-5327 (Toll Free); Fax: 301-470-1244; e-mail: edpubs@inet.ed.gov. Web site: http://www.ed.gov/offices/OUS/PES/eval.html.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Academic Standards; Board of Education Policy; Compensatory Education; *Educational Change; Educationally Disadvantaged; Elementary Education; Faculty Development; Mathematics Instruction; Parent Teacher Cooperation; *Poverty; Reading Instruction; School Policy; *Student Evaluation; Teaching Methods
IDENTIFIERS	*Elementary Secondary Education Act Title I; Reform Efforts; State Policy

ABSTRACT

The Longitudinal Evaluation of School Change and Performance (LESCP) analyses were organized around policies embodied in Title I of the Elementary and Secondary Education Act. This study measured changes in student performance in seventy-one high-poverty schools, which included many schools that had implemented standards-based reform policies. The study focused on: (1) analyzing student outcomes in reading and mathematics associated with specific practices in classroom curriculum and instruction; and (2) broadening its lens to learn about policy conditions, especially with regard to standards-based reform, under which the potentially effective classroom practices were likely to flourish. The report used data from standardized achievement tests, teacher surveys, district administrator and principal surveys, focus groups of school staff and parents, classroom observations, state and district policy statements, and student records. A particular asset of the study was its longitudinal database on more than 1,000 individual students, with student performance data linked to teachers' survey responses. This enabled researchers to pursue a detailed study of student performance in relation to the specific instructional conditions experienced by those students over time, with statistical controls for important background variables such as poverty. While confirming the harmful effects of poverty on the performance of students and schools, the findings point to some instructional practices that may boost performance in reading and mathematics. They also show an association between some of those practices and the standards-based reform approach of the current Title I program. The report contains four technical appendices. (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 457 306

PES

PLANNING AND EVALUATION SERVICE

THE LONGITUDINAL EVALUATION OF SCHOOL CHANGE AND PERFORMANCE (LESCP) IN TITLE I SCHOOLS

FINAL REPORT

VOLUME 2: TECHNICAL REPORT

2001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

U.S. DEPARTMENT OF EDUCATION ~ OFFICE OF THE DEPUTY SECRETARY
DOC #2001-20

UD034458

BEST COPY AVAILABLE

THE LONGITUDINAL EVALUATION OF SCHOOL CHANGE AND PERFORMANCE (LESCP) IN TITLE I SCHOOLS

FINAL REPORT

VOLUME 2: TECHNICAL REPORT

Prepared for:

U.S. Department of Education
Office of the Under Secretary
Contract No. EA 96008001

Prepared by:

Westat, Rockville, Md.
and
Policy Studies Associates, Washington, D.C.

2001

This report was prepared for the U.S. Department of Education under Contract No. EA96008001. The project monitor was Daphne Hardcastle in the Planning and Evaluation Service. The views expressed herein are those of the contractor. No official endorsement by the U.S. Department of Education is intended or should be inferred.

U.S. Department of Education
Rod Paige
Secretary

Office of the Deputy Secretary
William D. Hensen
Deputy Secretary

Planning and Evaluation Service
Alan L. Ginsburg
Director

Elementary and Secondary Education Division
Ricky Takai
Director

July 2001

This report is in the public domain. Authorization to produce it in whole or in part is granted. While permission to reprint this publication is not necessary, the citation should be the following: U.S. Department of Education, Office of the Deputy Secretary, Planning and Evaluation Service, *The Longitudinal Evaluation of School Change and Performance in Title I Schools, Volume 2: Technical Report*, Washington, D.C., 2001.

To order copies of this report, write

ED Pubs
Editorial Publications Center
U.S. Department of Education
P.O. Box 1398
Jessup, MD 20794-1398;

via fax, dial (301) 470-1244;

or via electronic mail, send your request to edpubs@inet.ed.gov.

You may also call toll-free 1-877-433-7827 (1-877-4-ED-PUBS). If 877 service is not yet available in your area, call 1-800-872-5327 (1-800-USA-LEARN); Those who use a telecommunications device for the deaf (TDD) or a teletypewriter (TTY), should call 8-800-437-0833.

To order online, point your Internet browser to www.ed.gov/pubs/edpubs.html.

This report is also available on the department's web site at www.ed.gov/offices/OUS/PES/eval.html.

On request, this publication is available in alternative formats, such as Braille, large print, audiotape, or computer diskette. For more information, please contact the Department's Alternate Format Center at (202) 260-9895 or (202) 205-8113.

CONTENTS

ACKNOWLEDGMENTS ix

INTRODUCTION: STUDY PURPOSES, DESIGN, AND SAMPLE CHARACTERISTICS 1

The Conceptual Model and How It was Implemented 2
LESCP Sample 4
Data Sources 5
Contents of This Report 7

OVERALL STUDENT PERFORMANCE ON TESTS 9

Standardized Tests 9
Available LESCP Test Scores 11
Cross-sectional Analyses 12
Longitudinal Sample 15
Poverty 17
Relationship Between the SAT-9 and State Assessments 20
Conclusions 22

CLASSROOM AND SCHOOL VARIABLES RELATED TO STUDENT PERFORMANCE 23

Variables and Methods Used in the Analysis of Student Performance 23
HLM Analysis and Results in Reading 28
HLM Analysis and Results in Mathematics 41
Summary 56

CONTEXT AND IMPLEMENTATION VARIABLES RELATED TO CLASSROOM AND SCHOOL-LEVEL PRACTICES 59

Poor and Initially Low-achieving Students' Access to Favorable Instructional Conditions 60
Policy Environment and Favorable Instructional Conditions 65
Summary 71

SUMMARY AND CONCLUSIONS 73

Summary of Findings 74
Conclusions 78

CONTENTS (CONTINUED)

Appendixes

A HIERARCHICAL LINEAR MODELING (HLM) ANALYSIS 79

B CHANGES IN THE TEACHER SURVEY ITEMS COMPRISING THE MEASURES OF STANDARDS-BASED REFORMS 101

C ADDITIONAL ANALYSES OF CHANGES 105

D RESULTS OF RELIABILITY ANALYSES FOR INDEX CONSTRUCTION 115

Tables

1-1 Summary of data collected 6

2-1 Grades tested, by year of data collection 11

2-2 LESCP sample sizes 11

2-3 Test taking rates for each year of the study 12

2-4 LESCP sample scores on the SAT-9 tests 13

2-5 National and urban norms for SAT-9 14

2-6 Sample size and mean scores for LESCP longitudinal sample 15

2-7 Difference in mean scores: LESCP longitudinal sample minus all LESCP test takers 16

2-8 Significant correlation coefficients among school rankings within the LESCP sample on the SAT-9 and on state assessment scores 21

3-1 Variables used to predict each longitudinal student's learning rate in the final HLM reading model 33

3-2 Variables used to predict each longitudinal student's third-grade score in the final HLM reading model 34

3-3 Final reading HLM model: Effects on score gains of longitudinal students, significant independent variables only 35

CONTENTS (CONTINUED)

Tables (continued)

3-4 Final reading HLM model: Effects on third-grade achievement of longitudinal students, significant independent variables only 37

3-5 Reading base-year scores and gains for longitudinal students with favorable instructional conditions, average longitudinal students, and national norms students 40

3-6 Variables used to predict each longitudinal student's learning rate in the final HLM mathematics model, control and student-level variables 47

3-7 Variables used to predict each longitudinal student's learning rate in the final HLM mathematics model, school-level variables 48

3-8 Variables used to predict each longitudinal student's third-grade score in the final HLM mathematics model 49

3-9 Final mathematics HLM model: Effects on score gains of longitudinal students, significant independent variables 50

3-10 Final mathematics HLM model: Effects on third-grade achievement of longitudinal students, significant independent variables only 53

3-11 Mathematics base-year scores and gains for longitudinal students with favorable instructional conditions, average longitudinal students, and national norms students 55

4-1 Direction of the main effects of teaching practices from the HLM analyses 61

4-2 Differences in favorable instructional indices for reading scores by student poverty, student achievement status in 1997, and school poverty concentration 62

4-3 Differences in favorable instructional indices for mathematics scores by student poverty, student achievement status in 1997, and school poverty concentration 66

4-4 Differences in favorable instructional indices for reading scores by policy environment 69

4-5 Differences in favorable instructional indices for mathematics scores by policy environment 70

CONTENTS (CONTINUED)

Tables (continued)

5-1 Overall results of the analysis of student performance in relation to very high levels (90th percentile) of specific instructional conditions 77

Figures

1-1 Conceptual framework 2

2-1 LESCP scores relative to national and urban norms for closed-ended reading 16

2-2 Average reading SAT-9 scale score for all LESCP students, grouped by school poverty level 18

2-3 Average mathematics SAT-9 scale score for all LESCP students, grouped by school poverty level 18

2-4 Average reading SAT-9 scale for all LESCP students, grouped by poverty 19

2-5 Average mathematics SAT-9 scale for all LESCP students, grouped by poverty 19

4-1 Conceptual framework 59

Boxes

3-1 Visibility of standards and assessments (reading) 29

3-2 Basic instruction in upper grades (reading) 30

3-3 Preparation for instruction (reading) 30

3-4 Rating of professional development (reading) 31

3-5 Outreach to low achievers' parents (reading) 31

3-6 Visibility of standards and assessments (mathematics) 42

3-7 Exploration in instruction (mathematics) 43

3-8 Presentation and practice in instruction (mathematics) 43

CONTENTS (CONTINUED)

Boxes (continued)

3-9 Preparation for instruction (mathematics) 44

3-10 Rating of professional development (mathematics) 45

3-11 Outreach to low achievers' parents (mathematics) 46

ACKNOWLEDGMENTS

We are indebted to many individuals whose contributions made the Longitudinal Evaluation of School Change and Performance (LESCP) possible. The project extended over 5 years and had many components and phases. Here we can only mention a few of the individuals who had a role in the design, conduct, and reporting of LESCP. We are especially grateful to our federal Project Officers Elois Scott and Daphne Hardcastle and to Audrey Pendleton and Jeffery Rodamar who served in an acting capacity. They provided invaluable substantive guidance, as well as support on the administrative and operational side. We wish to thank Alan Ginsburg, director of the Planning and Evaluation Service (PES), whose ideas and questions helped us formulate the research design, study questions, and analyses for LESCP. Ricky Takai, director of the Elementary and Secondary Division of PES, asked the hard questions as the study and analyses progressed and kept us on course. Other staff of PES who contributed in various capacities over the life of LESCP are Stephanie Stullich, Joanne Bogart, and Barbara Coates. From the Compensatory Education Programs office, Mary Jean LeTendre and Susan Wilhelm provided thoughtful advice. We also thank Valena Plisko and Michael Ross of the National Center for Education Statistics for their helpful expertise.

Members of the LESCP Technical Work Group were active contributors to the study effort from the very beginning. They provided advice on the conceptual framework and research design, reviewed and advised on plans for analysis, and ultimately on the analysis itself. We wish especially to thank them for their reviews and thoughtful comments and recommendations during the development of this report. Members of the Technical Work Group are identified below. Each member served for the entire study period except where noted.

Technical Work Group Members

Dr. David Cordray, Department of Psychology and Human Development, Vanderbilt University

Dr. Judith McDonald, (1998–2000) Team Leader, School Support/Title I, Indian Education, Oklahoma State Department of Education

Dr. Andrew Porter, Wisconsin Center for Education Research, School of Education, University of Wisconsin-Madison

Dr. Margaret Goertz, Consortium for Policy Research in Education

Dr. Mary Ann Millsap, Vice President, Abt Associates

Dr. Jim Simmons, Program Evaluator for the Office of Student Academic Education, Mississippi Department of Education

Dr. Joseph F. Johnson, Charles A. Dana Center, University of Texas at Austin

Ms. Virginia Plunkett, (1997–98) Colorado Department of Education

Dr. Dorothy Strickland, Graduate School of Education Rutgers University

A large number of Westat and subcontractor staff contributed to LESCP. Linda LeBlanc of Westat served as project director and Brenda Turnbull of Policy Studies Associates (PSA) as co-project director. From Westat, Alexander Ratnofsky, Patricia Troppe, William Davis, Ann Webber, and Camilla Heid served on the analysis and reporting team. Raymond Olsen, Stephanie Huang, Bahn Cheah, Alan Atkins, Sandra Daley, and Cyril Ravindra provided systems support for analysis and survey operations. Juanita Lucas-McLean, Therese Koraganie, and Dawn Thomas handled all survey data collection and processing. Heather Steadman provided secretarial support, Carolyn Gatling provided word processing support, and Arna Lane edited this report.

At PSA, Megan Welsh played a major role in the study's analysis and reporting. Other PSA staff members who also made substantial contributions to the study are Joelle Gruber, Ellen Pechman, Ullik Rouk, Christina Russell, and Jessica Wodatch.

Jane Hannaway of the Urban Institute coordinated the teams of site visitors from the Urban Institute and along with Nancy Sharkey assisted with analyses of some of the early study data.

Everett Barnes and Allen Schenck organized and oversaw the work of RMC Research Corporation staff from that company's Portsmouth, Denver, and Portland offices over three cycles of site visits and data collection for LESCP.

The analyses of student achievement benefited from the help of Aline Sayer, of the Radcliffe Institute for Advanced Study, Harvard University, who lent her expertise in the statistical technique of hierarchical linear modeling.

The Longitudinal Evaluation of School Change and Performance would not have been possible without the support and participation of school principals and teachers who welcomed us into their schools and provided the heart of the information on which this report is based. We are particularly indebted to the school administrators who took on the task of coordinating all LESCP activities at their schools. Our greatest thanks go to the 12,000 students who took part in the study and did their best on the assessments we administered each spring and to their parents for allowing their children to contribute to the body of knowledge on school reform.

INTRODUCTION: STUDY PURPOSES, DESIGN, AND SAMPLE CHARACTERISTICS

The Longitudinal Evaluation of School Change and Performance (LESCP) design and analyses were organized around the policies embodied in Title I of the Elementary and Secondary Education Act, as amended in 1994. Since its original enactment in 1965, Title I has been intended to improve the learning of children in high-poverty schools, with a particular focus on those children whose previous achievement has been low. Therefore, this study measured changes in student performance in a sample of Title I schools, and its analyses included a special look at those students with initially low achievement. The study was in the tradition of past work addressing school practices and policies that can contribute to higher achievement in Title 1 schools.[1] The study had a dual focus: (1) analyzing the student outcomes associated with specific practices in classroom curriculum and instruction; and then (2) broadening its lens to learn about the policy conditions—especially with regard to standards-based reform—under which the potentially effective classroom practices were likely to flourish.

The second focus considers the provisions of Title I enacted in 1994 that strongly encourage states, school districts, and schools to pursue a standards-based approach to educational improvement. The standards-based approach relies on aligned frameworks of standards, curriculum, student assessment, and teacher professional development to set clear goals for student performance and to help organize school resources around those goals. It is an approach that several states and some large districts began to put in place earlier in the 1990s. Several large federal programs, prominently including Title I, adopted the philosophy of standards-based reform in 1994.

This chapter describes the conceptual model used for the study's data collection and analysis and how the model was implemented for the study. It then describes the variation found in the study's purposive sample with regard to standards-based reform policies and school characteristics. The final section of the chapter highlights the major data sources for the study.

---

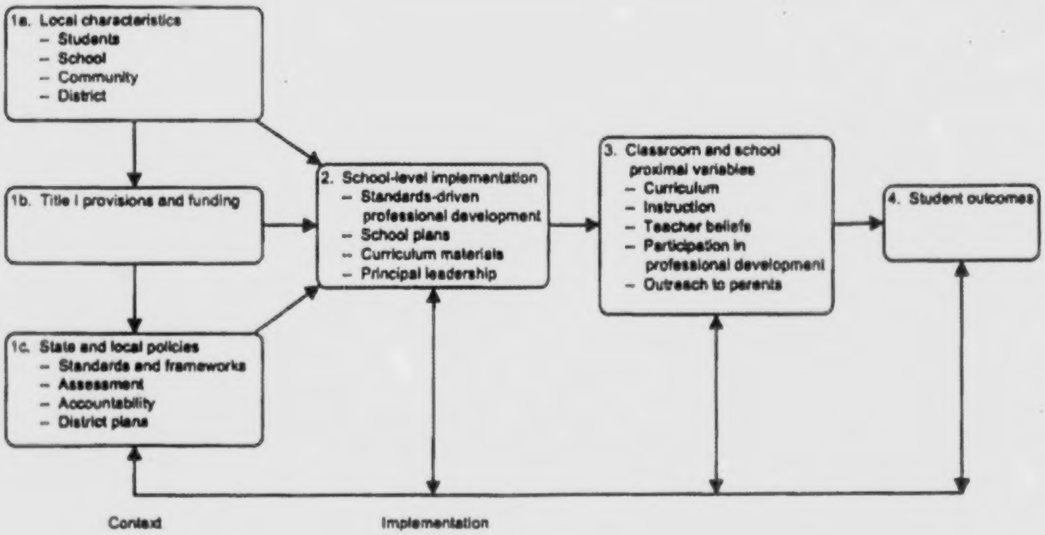
[1] See, for example, D.R. Lee, R.A. Carriere, A.H. MacQueen, L.H. Poynor, and M.S. Rogers. (1981). Successful practices in high-poverty schools, (Technical Report No. 16) in the *Study of the Sustaining Effects of Compensatory Education on Basic Skills*. Santa Monica, Calif.: System Development Corporation; M.M. Kennedy, B.F. Birman, and R.E. Demaline. (1986). *The effectiveness of Chapter I services*. Washington, D.C.: U.S. Department of Education; M.S. Knapp, P.M. Shields, and B.J. Turnbull. (1992). *Academic challenge for the children of poverty*. Washington, D.C.: U.S. Department of Education.

The Conceptual Model and How It was Implemented

The study's conceptual framework, depicted in Figure 1-1, shows the study's design for tracing the complicated path by which policy might affect student performance. Beginning on the right of the framework, Box 4 in Figure 1-1 represents the student-level goal of Title I and other policies, improved student achievement. In this study, most of our analyses used the Stanford Achievement Test, Ninth Edition (SAT-9) tests of reading and mathematics as measures of student achievement. Students took the SAT-9 tests in the third and fourth grades in 1997, the fourth grade in 1998, and the fourth and fifth grades in 1999; this permitted us to track the performance gains of individual students over time and also the performance of successive cohorts of fourth graders. We discuss the specifics of these measures, including pros and cons, in Chapter 2 of the report. For purposes of the conceptual framework, we note merely that some of our analyses focused on the scale scores attained by students in a particular grade, and others focused on the score gains made by individual students over 2 years (from third grade to fifth grade).

Box 3 in Figure 1-1 represents *proximal* variables, those that might plausibly exert a direct influence on student achievement in reading and mathematics. They include classroom curriculum (what was taught) and instruction (how the material was taught). For kindergarten to fifth-grade classrooms in

Figure 1-1. Conceptual framework


1a. Local characteristics
– Students
– School
– Community
– District
1b. Title I provisions and funding
1c. State and local policies
– Standards and frameworks
– Assessment
– Accountability
– District plans
2. School-level implementation
– Standards-driven professional development
– School plans
– Curriculum materials
– Principal leadership
3. Classroom and school proximal variables
– Curriculum
– Instruction
– Teacher beliefs
– Participation in professional development
– Outreach to parents
4. Student outcomes
Context
Implementation

BEST COPY AVAILABLE

schools participating in this study, teachers responded to questionnaires about their curriculum and instruction in reading and mathematics in all three study years. They also answered questions about their beliefs, the professional development they received, and their outreach to parents. In addition to measures for individual teachers, Box 3 also includes measures of curriculum, instruction, and instructional support—like professional development—averaged at the school level. This is because instructional influences on students may come from features of the whole school environment, not just one classroom.

Demographic conditions, especially individual and school poverty, can be important influences on student achievement. Thus, the analyses that explored the connections between Box 3 and Box 4 in Figure 1-1 paid attention to control variables such as poverty and school size.

At the same time that the study viewed the Box 3 variables (in Figure 1-1) as proximal inputs to student achievement (Box 4 in Figure 1-1), it also treated them as outcome measures. Logically, one would expect curriculum, instruction, and instructional supports to reflect a combination of influences that would include policies from the local, state, and federal levels (Boxes 1b and 1c in Figure 1-1) and socioeconomic conditions impinging on the school (Box 1a in Figure 1-1). These influences would be filtered through a school's implementation choices, which are represented by Box 2 in Figure 1-1—for example, what professional development the school offered, and how vigorously the principal sought to impress the importance of outside standards on the teachers. The surveys developed for this study permitted us to understand Box 2 from the perspective of all the teachers in the school, using our measures of the extent to which teachers had actually participated in professional development or reported that they were familiar with standards. While these measures give an indirect window on what the school did to promote implementation of outside policies, we would argue that it was an important window, showing what messages from the school were received by the teachers.

Looking across Figure 1-1, readers will notice that policies appear not as direct influences on student performance but as influences that are mediated through school-level policy implementation and teacher-level practices. This reflects our conviction that a policy enacted in Washington, D.C.; a state capital; or a school district's central office cannot possibly affect student performance unless and until schools and teachers do something different. Accordingly, as just described, the LESCP-analysis-built models of student performance that tested the possible influence of many variables drawn from Box 3 in Figure 1-1. Having done that, we then treated Box 3 variables as outcomes and explored the extent to which important Box 3 variables were found in environments of standards-based reform.

LESCP Sample

This study looked in depth at a purposive sample of state and local policy environments rather than employing a larger, nationally representative sample of Title I schools. To assess school and classroom responses to standards-based reform, the study focused on states and districts that had enacted standards-based reform some years earlier.[2] The states varied in their approach to reform—for example, some put high-stakes assessment in place early on, while others began their reforms with a process of developing content standards. Of the seven states in the sample, five were arguably embarked on some version of standards-based reform in 1996 when the sample was drawn. Although the other two states were doing less with standards-based reform in 1996, they moved in that direction over the course of the study, and one of them moved quite rapidly. This left the study with somewhat less variation at the state level than was originally expected. The 18 participating districts presented a similar pattern: none was untouched by standards-based reform, although they varied in the alacrity and thoroughness with which they enacted each of several kinds of standards-based policies, both in response to state requirements and on their own initiative. In short, the LESCP schools were subject to some variation in the kinds of policies enacted by their states and districts—but it is important to recognize that all were subject to some policy activity in standards, assessment, or accountability.

To better understand the policy environments that schools were operating in, we used documents provided by the districts' offices to create ratings for each of the 18 districts on several indicators of standards-based reform policies in 1998. We focused this analysis at the district level to capture both state and district policy. Taking this approach was necessary because our sample deliberately included pairs of states and districts that had initially taken different policy stances on standards-based reform. The sample included, for example, districts that state officials described as reluctant to implement an aggressive standards-based agenda initiated at the state level. It also included districts that had independently established their own standards-based framework in the absence of such a framework statewide.

The 71 schools in the sample all received Title I funds, and most had very high levels of poverty. Of the 71 schools, 59 were operating schoolwide programs in 1998-99 (up from 58 in 1997-98 and 54 in 1996–97). This reflected, in part, the high levels of poverty in participating schools: 15 schools had more than 90 percent of their students living in poverty, 25 schools had between 75 percent and 90 percent, 21 schools between 50 and 75 percent, and 10 schools had fewer than 50 percent. In all schools, the poverty rate was higher than 35 percent.

---

[2] The study could not meet its mandate simply by looking at schools' responses to the provisions of the 1994 law because the timeline for the study does not mesh well with the timeline for the law's implementation. For example, Title I does not require full implementation of standards-based accountability for schools until the school year 2000–01, 2 years after LESCP data collection ends. Thus, only in those states and districts that had already enacted standards-based reform some years ago—before the Title I provisions were enacted—would it be possible to expect widespread, classroom-level effects as early as the LESCP data collection period.

Data Sources

This report is based on three rounds of data gathered in spring 1997, spring 1998, and spring 1999 from students and school staff. The LESCP study collected repeated measures of students' performance, teachers' reported behavior and opinions, and the school's policy environment in 71 schools. These schools, all of which receive funds under Title I, were nested in a purposively selected sample of 18 districts in 7 states. The schools were not statistically representative of high-poverty schools in the nation as a whole, in their states, or even in their districts. However, the study provided a rich database that permitted the analysis of differences across students, classrooms, schools, and policy environments at any one time and also across school years. A summary of the data collected is shown in Table 1-1.

This report draws most heavily on three of the study's data sources:

- The tests administered to students who were in the third grade in 1997, fourth grade in 1998, and fifth grade in 1999;
- Surveys completed by teachers in each year regarding topics that include their classroom curriculum and instruction in reading and mathematics, their knowledge and instruction with regard to standards-based reform and instruction, their professional development over the past 12 months, and their outreach to parents of low-achieving students; and
- Documents collected from school districts regarding policies related to standards based reform.

Data on schools' performance on state or local assessments were also collected and used. First, we analyzed the relationship between school performance on the state assessments and on the SAT-9, relative to the rest of the sample schools in that state. We also looked at the extent to which the study's proximal variables, averaged at the school level, were related to trends over time in schools' performance on their state tests.

Table 1-1. Summary of data collected

Instrument	Number of respondents	Format	Information domains
Survey of teachers in schoolwide schools	All classroom teachers in sampled schools (approximately 20 teachers per school)	Mailed, with on-site collection	Curriculum, instructional practice, knowledge of standards, assessment methods, professional development, parental involvement
Survey of Title I teachers in Title I targeted assistance schools	All Title I teachers in sampled schools (approximately 6 teachers per school)	Mailed, with on-site collection	Curriculum, instructional practice, knowledge of standards, assessment methods, professional development, parental involvement, Title I services
Survey of non-Title I teachers in Title I targeted assistance schools	All classroom teachers in sampled schools (approximately 20 teachers per school)	Mailed, with on-site collection	Curriculum, instructional practice, knowledge of standards, assessment methods, professional development, parental involvement
Title I District Administrator Interview Guide	18 (one per district)	On-site interview	School improvement process, capacity building, parental involvement
Principal Interview Guide (different instrument for schoolwide and targeted assistance schools)	71 (one per school)	On-site interview	Planning, coordination, and delivery of Title I services; curriculum and instruction; measuring and improving student performance; parental involvement; capacity building.
Guide for school staff focus group	2 per school; each with 6–10 participants	On-site focus group session	Title I implementation, coordination of curricula, school-level capacity, and parental involvement
Guide for parent focus group	1 per school; 6–10 participants	On-site focus group session	Parental involvement, parent resources
Classroom observation protocol	4 lessons per school, 2 selected at each of the tested grades	On-site observation	Verify and expand on content and pedagogy information from the teacher survey
Document collection and review guide	71 schools	On-site at school and district	School-level and district-level documents, data, and materials
Student information form	Average 67 students per grade per school	On-site data abstraction	Student-level demographics and program participation
Standardized achievement tests	Average 67 students per tested grade per school	SAT-9, plus other tests on file	Reading and mathematics test scores

Contents of This Report

This report essentially works its way from right to left in the conceptual model. The next chapter describes the achievement of students in the sample, both the full sample and the smaller group of students who were followed over a 2-year period as they moved from third grade through fifth grade. In Chapter 3, we present the results of the analysis of influences on student achievement: what instructional conditions were associated with higher levels of student performance, either initially or over time. We use those findings in Chapter 4 to identify the extent to which students had access to favorable instructional conditions—the variables that seemed to matter for achievement—if they lived in poverty or if their state or district had enacted particular dimensions of standards-based reform. Conclusions appear in Chapter 5.

This report has four technical appendixes. In Appendix A, we describe in detail the statistical approach we used in examining the relationships among student achievement, student and school characteristics, and classroom instructional practices. Appendix B consists of tables that show how the various measures of instructional practices that we used changed over the 3 years of data collection. We performed some secondary analyses of factors related to student test scores. The results of these analyses are reported in Appendix C. In Appendix D, we present the results of the reliability analyses for the indices we constructed to measure instructional practices.

OVERALL STUDENT PERFORMANCE ON TESTS

The student-level goal of Title I and other policies is improved student achievement (see Box 4 in Figure 1-1). A major source of student achievement data for this study was the student testing conducted with third and fourth graders in spring 1997, with fourth graders in spring 1998, and with fourth and fifth graders in spring 1999. This chapter describes the standardized tests; the students who took each of the tests in spring 1997, 1998, and 1999 (as well as the extent and causes of missing data); overall results for all the students and for the subset of students who were tested in all 3 years; and a comparison of performance at the school level between the standardized tests and states' own assessments.

This chapter describes student performance as background to an investigation in Chapter 3 of the relationship between the proximal variables associated with classroom and school instructional practices and student performance. The data highlight the fact that, on average, students in the Longitudinal Evaluation of School Change and Performance (LESCP) schools underperform in reading and in mathematics in the third grade when compared with national norms and that, on average, do not close the gap by the fifth grade. The data also show strong correlations between both student-level and school-level poverty and student achievement. This sets the context for the Chapter 3 analyses that seek to identify classroom and school-level practices that work to overcome the effects of poverty and poor performance in the early grades.

Standardized Tests

The study administered norm-referenced achievement tests in reading and mathematics, the Stanford Achievement Test, Ninth Edition (SAT-9), to participating students. According to the publisher, the closed-ended mathematics test aligns with the National Council of Teachers of Mathematics standards in effect during the LESCP field period,[1] and one section of the closed-ended reading test (the reading comprehension subtest) aligns with the National Assessment of Educational Progress.[2] Separate scores were obtained for each of the four tests in spring 1997, 1998, and 1999:

---

[1] *Stanford 9 Special Report: Mathematics*. (1997). San Antonio, Tex.: Harcourt Educational Measurement.

- Overall closed-ended reading;
- Open-ended reading;
- Overall closed-ended mathematics; and
- Open-ended mathematics.

The closed-ended reading test is composed of two subtests, vocabulary and comprehension, at the grades administered in the LESCP. The vocabulary subtest assesses vocabulary knowledge and skills with synonyms, context clues, and multiple word meanings. The reading comprehension subtest uses a reading selection followed by multiple choice questions to measure modes of comprehension (initial understanding, interpretation, critical analysis, and process strategies) within the framework of recreational, textual, and functional reading. The open-ended reading test contains a narrative reading selection in the recreational reading content cluster followed by nine open-ended questions that measure initial understanding, interpretation, and critical analysis.

The closed-ended mathematics test is composed of problem-solving and procedures subtests. Five processes are assessed in the problem-solving subtest: problem solving, reasoning, communication, connections, and thinking skills. Concepts of whole numbers, number sense and numeration, geometry and spatial sense, measurement, statistics and probability, fraction and decimal concepts, patterns and relationships, estimation, and problem-solving strategies are measured. The procedures subtest covers number facts, computation using symbolic notation, computation in context, rounding, and thinking skills. The open-ended mathematics assessment presents nine questions or tasks around a single theme. Ability to communicate and reason mathematically and to apply problem-solving strategies are assessed. The content clusters for the open-ended mathematics test are number concepts, patterns and relationships, and concepts of space and shape.

The number of students for whom we have test scores varies by the test because not every district had its students take each component test. Both the mathematics and reading open-ended tests included all districts in the LESCP study. However, one district did not participate in one component of the closed-ended reading test, and another district's scores in closed-ended reading were converted from SAT-8 scores to SAT-9 scores using equating methods suggested by the test publisher.

---

[2] *Stanford 9 Special Report: Reading*. (1997). San Antonio, Tex.: Harcourt Brace Educational Measurement.

Available LESCP Test Scores

In this report, we analyze LESCP test scores from data collected during spring 1997, 1998, and 1999. Table 2-1 shows the basic sources of data studied here. We had scores for the third and fourth grades in 1997, the fourth grade in 1998, and the fourth and fifth grades in 1999. We paid particular attention in the analysis to the cohort of students who were third graders in spring 1997. Because we had repeated measurements on many of these students, we could measure score growth with a reliable baseline score.

Table 2-1. Grades tested, by year of data collection

Year of data collection		
1997	1998	1999
Third grade	No testing conducted	No testing conducted
Fourth grade	Fourth grade	Fourth grade
No testing conducted	No testing conducted	Fifth grade

Table 2-2 shows the total number of third-, fourth-, and fifth-grade LESCP students tested for each of the four tests in spring 1997, 1998, and 1999. The minimum number of students for any test, grade, and year is 2,567. This is an appreciable sample and should allow us to make reliable conclusions.[3]

Table 2-2. LESCP sample sizes

	Third grade	Fourth grade			Fifth grade
Test	1997	1997	1998	1999	1999
Reading closed-ended	2,813	2,692	2,567	3,213	3,311
Reading open-ended	3,646	3,535	3,438	3,503	3,328
Mathematics closed-ended	3,226	3,073	2,987	3,052	2,871
Mathematics open-ended	3,723	3,503	3,400	3,455	3,326

---

[3] By comparison, the national percentile and mean estimates developed by the SAT-9 publisher are based on samples size of 4,000 to 5,000.

Table 2-3 documents the reasons why some students did not take the closed-ended tests. Similar percentages were found for nontest takers for the open-ended tests. In 1997 and 1998, roughly 10 percent of the parents exercised their privilege to excuse their children from participating in the study. This percentage increased to 14 percent in 1999. Approximately 5 percent of the students enrolled in the appropriate grades in the LESCP sample of schools were excused by the school from taking the test because of their determination that the students' limited English or disability made test taking inappropriate. Schools were asked to use the criteria they used for state or districtwide testing to determine who should be excused for either of these reasons. Another 3 percent were absent the day of the test, and the test scores for yet another 3 percent or 4 percent were not available for other reasons.

Table 2-3. Test taking rates for each year of the study

	Closed-ended reading			Closed-ended mathematics		
	1997	1998	1999	1997	1998	1999
Took the test	77%	80%	77%	78%	79%	76%
Parent refused	9%	10%	14%	9%	10%	14%
Limited-English proficiency (LEP) or disabled	6%	4%	5%	6%	4%	5%
Absent day of test	3%	3%	3%	3%	3%	3%
Other	4%	3%	2%	3%	4%	3%
TOTAL	100%	100%	100%	100%	100%	100%

In addition to focusing our analyses on the cohort of students who were third graders in spring 1997, we identified a subset of these students for further analysis. These students, who were tested in all 3 years, were called the *longitudinal sample*. In contrast to the population of all students in the cohort, this group was less mobile and may have enjoyed other advantages. We discuss the characteristics of these students in Section 2.4.

Cross-sectional Analyses

We compared the performance of LESCP students with national and urban reference groups and with proficiency levels identified by the test publisher. On average, students in the LESCP sample of schools scored below national norms and urban norms in all years and grades tested. Table 2-4 shows cross-sectional data on test performance for the entire LESCP sample for 1997, 1998, and 1999. The data are shown in several forms: overall mean scores, the national percentile and grade-equivalent that these

Table 2-4. LESCP sample scores on the SAT-9 tests

Grade (Year)	Test	Mean score	National percentile of mean score	Grade[1] equivalent of mean score	% Level 1: below satisfactory	% Level 2: partial mastery	% Level 3: solid performance	% Level 4: superior performance
Third Grade (1997)	Reading closed-ended	602	38	3.4	32%	40%	24%	4%
	Reading open-ended	575	37	3.2	37%	32%	18%	12%
	Mathematics closed-ended	592	43	3.5	30%	46%	21%	4%
	Mathematics open-ended	582	32	2.9	29%	36%	25%	9%
Fourth Grade (1997)	Reading closed-ended	623	35	4.1	35%	39%	19%	7%
	Reading open-ended	598	39	4.2	36%	41%	17%	5%
	Mathematics closed-ended	614	39	4.4	33%	39%	22%	6%
	Mathematics open-ended	590	23	3.5	43%	38%	14%	6%
Fourth Grade (1998)	Reading closed-ended	621	34	4.0	34%	38%	21%	8%
	Reading open-ended	602	42	4.3	31%	44%	20%	5%
	Mathematics closed-ended	614	39	4.4	34%	39%	21%	6%
	Mathematics open-ended	590	23	3.5	47%	34%	13%	6%
Fourth Grade (1999)	Reading closed-ended	623	35	4.0	36%	39%	19%	6%
	Reading open-ended	595	37	4.0	40%	40%	16%	4%
	Mathematics closed-ended	615	40	4.4	38%	38%	20%	4%
	Mathematics open-ended	591	24	3.6	47%	35%	13%	5%
Fifth Grade (1999)	Reading closed-ended	640	36	5.0	31%	48%	19%	2%
	Reading open-ended	606	32	4.6	52%	31%	13%	4%
	Mathematics closed-ended	636	38	5.4	47%	33%	16%	4%
	Mathematics open-ended	612	28	4.9	41%	38%	15%	6%

[1] The tests are normed for spring of each grade. For example, the norm grade equivalent for the third grade is approximately 3.8, for fourth grade 4.8, and for fifth grade 5.8.

BEST COPY AVAILABLE

means represent, and the percentage of LESCP test takers who performed at particular competency levels on each test in each year. These levels are described as corresponding to the kinds of performance levels that Title I encourages for state assessment data (e.g., "excellent," "proficient," and the like) and set by the SAT-9 test publisher. (See *Technical Data Report* for the SAT-9, 1997, Harcourt Brace & Company).

For comparison, Table 2-5 shows the national and urban norms by test. National and urban norms were taken from the SAT series (1996) based on a representative sample of student scores in spring 1995. Because urban means were not available for the open-ended test, urban medians were used. On all tests, the LESCP students fell below the national norms by 4 points to 23 points.[4] Additionally, the LESCP students scored below the urban norm on all tests.

Table 2-5. National and urban norms for SAT-9

Test	Third grade		Fourth grade		Fifth grade	
	National mean	Urban median	National mean	Urban median	National mean	Urban median
Reading closed-ended	614	607	637	634	654	647
Reading open-ended	586	579	606	609	629	624
Mathematics closed-ended	600	593	624	624	646	639
Mathematics open-ended	602	590	612	609	626	621

On three of the four tests, the means and distributions of test scores for the LESCP sample of fourth graders remained essentially indistinguishable from year to year. On the open-ended reading test, where there was a statistically significant rise in 1998, there was a subsequent decline in 1999.

We note that the cross-sectional results for fourth graders were obtained on three different fourth-grade classes in the LESCP schools. In the next section, we analyze the change within the cohort of students who were third graders in spring 1997. In this group, we can more accurately determine whether any statistically significant changes were due to the educational experiences of participating students during the fourth and fifth grades.

---

[4] The conclusions based on the differences from national and urban norms are valid if the national and urban scores have not changed substantially since the norming year, 1995.

Longitudinal Sample

For the longitudinal sample, we had a reliable baseline score and thus we could accurately assess the score gain made between spring 1997 and spring 1999. In comparison to using all test takers for this analysis, the use of the longitudinal sample for analysis has the following two disadvantages: it reduces the sample size; and it limits the generalizability of the conclusions to those students who spent third, fourth, and fifth grade at the same school. As shown below, the students in the longitudinal sample scored higher on standardized tests, on average, than did all students in a grade. However, the advantage of using the longitudinal sample is that it allows us to conduct a longitudinal study; to distinguish changes over time within students from differences among students in their baseline levels. We believe that the advantages of basing conclusions on the longitudinal sample outweigh the disadvantages.

LESCP students were tested on grade. If a student progressed from third grade to fourth grade to fifth grade, we required that the student take the third-grade form of the SAT-9 test in 1997, the fourth-grade form in 1998, and the fifth-grade form in 1999, to be included in the longitudinal sample.

Table 2-6 shows the number of students in the longitudinal sample and the mean scores for this group for the four tests taken in spring 1997, 1998, and 1999. The ratios of longitudinal test takers to third-grade test takers in 1997 ranged from 42 percent in closed-ended mathematics to 50 percent in closed-ended reading. That is, between 40 percent and 50 percent of those third graders tested in 1997 were tested again as fourth graders in 1998 and fifth graders in 1999.

Table 2-6. Sample size and mean scores for LESCP longitudinal sample

Test	Sample size	Mean 1997	Mean 1998	Mean 1999
Reading closed-ended	1,401	607	628	646
Reading open-ended	1,656	581	607	612
Mathematics closed-ended	1,358	597	621	642
Mathematics open-ended	1,642	586	593	617

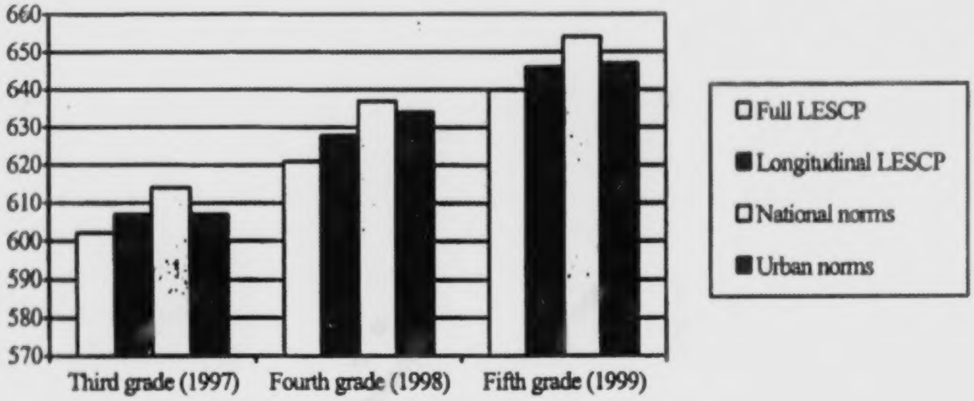
Table 2-7 shows the difference in mean scores between the LESCP longitudinal sample and all the LESCP students in the cohort. As noted above, longitudinal students scored higher than the other test takers in all four tests in all 3 years. The difference in means ranged from 3 to 7 points.

Table 2-7. Difference in mean scores: LESCP longitudinal sample minus all LESCP test takers

Test	1997	1998	1999
Reading closed-ended	5	7	6
Reading open-ended	6	5	6
Mathematics closed-ended	5	7	6
Mathematics open-ended	4	3	5

A pictorial display of the relationships among average closed-ended reading scores for all LESCP students, longitudinal LESCP students, national norms, and urban norms is shown in Figure 2-1. Scores for the other tests show essentially similar patterns.

Figure 2-1. LESCP scores relative to national and urban norms for closed-ended reading



The LESCP longitudinal sample students stayed at the same level of performance relative to national norms between third and fifth grade on both closed-ended tests. That is, the number of points gained by the longitudinal sample was similar to the difference in points between the third and fifth grade norm groups.[5] For example, the increase in score between the third and fifth grades for the longitudinal sample in closed-ended reading was 39 points, while the difference on medial national norms for the third and fifth grades was 40 points. The only difference between the national norm group and the longitudinal sample was found on the open-ended reading test.[6] On this test, the longitudinal sample gained 12 points less than the norm group.

---

[5] Note that the test publishers did not follow a group of students over time and track their progress as part of their norming procedure. All norm groups were tested at the same time; norm group comparisons across grade levels may not accurately estimate normal growth over time.

[6] The open-ended tests were normed with a smaller sample than the closed-ended tests, this may account for these differences.

Poverty[7]

Year by year, student performance revealed the disadvantage of attending a high-poverty school or experiencing family poverty.[8] In both reading and mathematics, the schools serving higher proportions of poor students started out with lower achievement (as measured by third-grade score) and failed to close the gap over time—although at least it can be said that the gap did not widen. Figures 2-2 and 2-3 display the scores on the closed-ended tests of reading and mathematics, by year and level of school poverty, for all LESCP test takers in third grade in 1997, fourth grade in 1998, and fifth grade in 1999. The schools are divided into four categories by percentage of students eligible for free or reduced-price lunch: (1) the least-poor group of schools, where fewer than 50 percent of the students were eligible, (2) 50 percent to 75 percent, (3) 75 percent to 90 percent, and (4) the poorest group, with more than 90 percent eligible. In schools where fewer than 50 percent of the students were eligible, average student performance was above that of the national norm group in both reading and mathematics. At the other end of the poverty scale, students in the highest poverty schools started out scoring lowest on the SAT-9 and continued to score the worst over time, maintaining an equally large gap from the national norm group between third and fifth grade.

The combined effects of family poverty and attending a high-poverty school were still more serious, as Figures 2-4 and 2-5 display. In both reading and mathematics, we found that students in higher poverty schools had lower scores. At each level of school poverty, those students who were eligible for free or reduced-price lunch had lower scores. For example, in third-grade reading, the group of students with the lowest average score was the group of students eligible for free or reduced-price lunch and attended the highest poverty schools. The group that scored highest was the group of students not eligible for free or reduced-price lunch and attended the lowest poverty schools. Thus, there was a compounding negative effect of being poor and in a high-poverty school.

---

[7] The LESCP's sole measure of poverty is eligibility for free or reduced-price lunch. This is a measure of poverty that is commonly used in Title I. For example, for a school to be able to implement its Title I program as a schoolwide program, the school must have at least 50 percent of its students eligible for free or reduced-price lunch. However, this measure covers a broad range of family incomes and poverty conditions. Other, more precise, measures of poverty may yield different findings.

[8] LESCP data also confirmed a very high correlation between race and poverty in the sampled schools. African American and Hispanic students were significantly more likely to be eligible for free or reduced-price lunch than white students.

Figure 2-2. Average reading SAT-9 scale score for all LESCP students, grouped by school poverty level

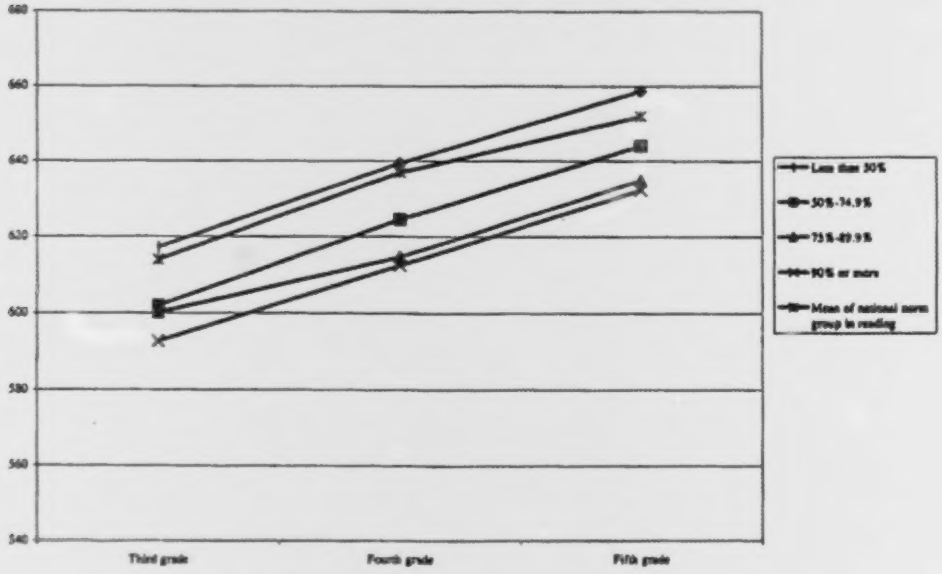
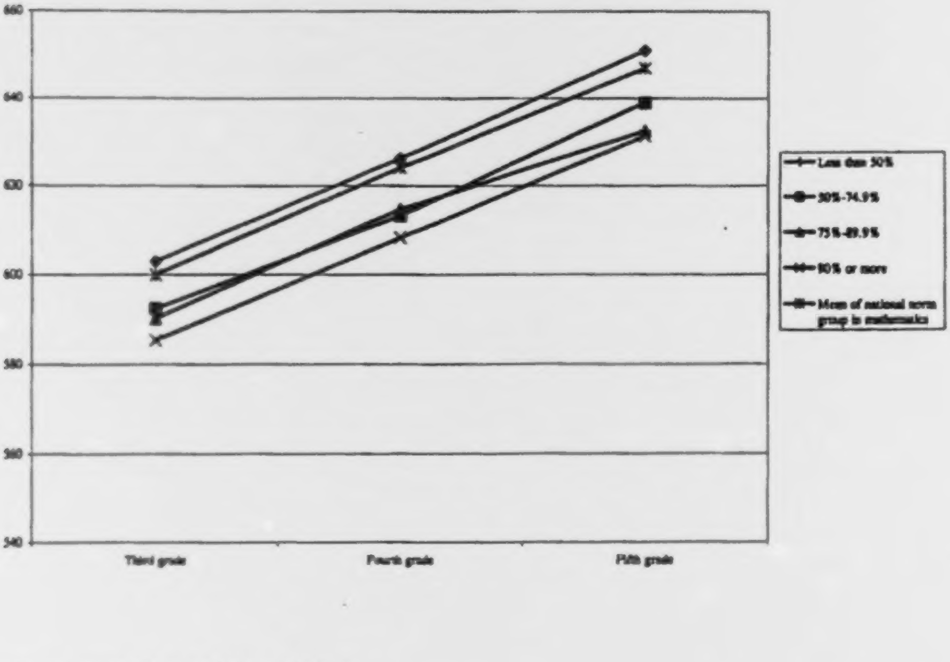


Figure 2-3. Average mathematics SAT-9 scale score for all LESCP students, grouped by school poverty level



BEST COPY AVAILABLE

Figure 2-4. Average reading SAT-9 scale for all LESCP students, grouped by poverty

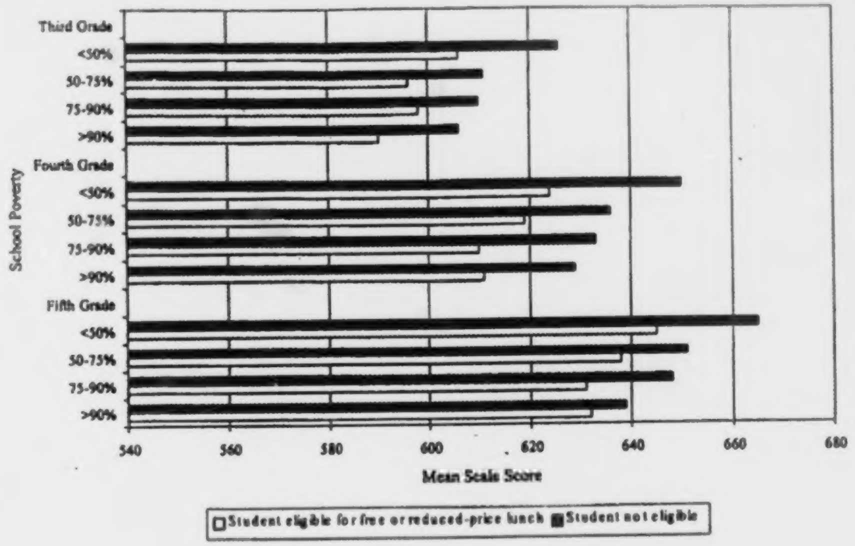
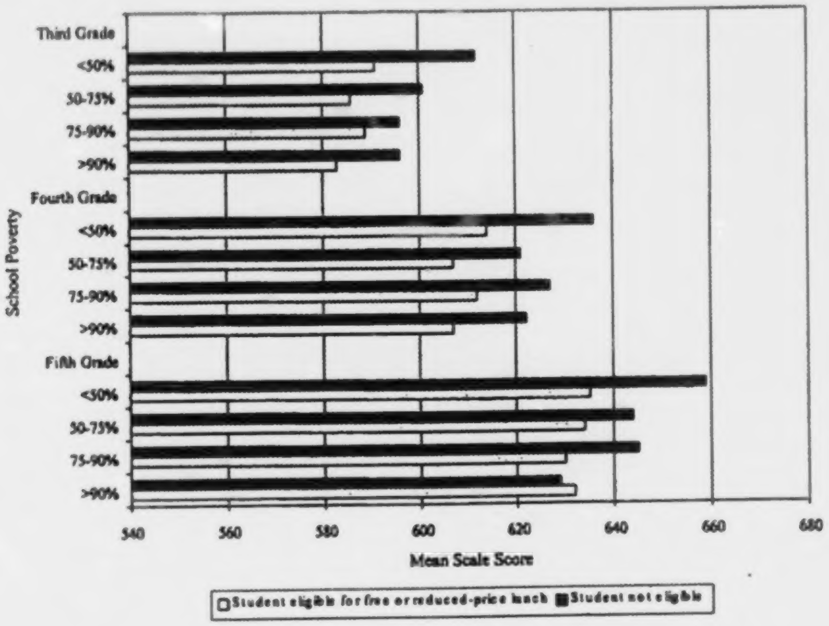


Figure 2-5. Average mathematics SAT-9 scale for all LESCP students, grouped by poverty



These relationships between poverty and achievement illustrate the problem that Title I addresses. Studies like this one assess the extent to which policies in Title I, such as the program's adoption of standards-based reform, are making the kind of changes in the classroom that might alleviate the achievement deficits wrought by poverty. The rest of this report furnishes our answers. But first we briefly examine the relationship between SAT-9 achievement scores (the primary outcome measures for LESCP) and state assessments (the primary outcome measures used by states and school districts).

Relationship Between the SAT-9 and State Assessments

The analyses conducted for this report, focusing on the student outcomes associated with particular aspects of classroom curriculum and instruction, emphasized trends in individual student performance across years. Student performance on the SAT-9 gave us performance data in a common metric across all the study's classrooms, with individual test scores that could be associated with students' individual demographic characteristics, and their own teachers' survey responses concerning curricular and instructional attitudes and practices. We did not have a comparable level of detail regarding performance on state tests, and of course those tests varied across states. However, Title I charges states and school districts with improving performance in relation to state standards, as measured by state assessments. Therefore, it was worth checking how well the SAT-9 results match up with the results of state assessments.

Comparisons were possible when states administered an assessment at the same grade level and in the same year as the LESCP study administered the SAT-9 (third grade in 1997, fourth grade in any of the 3 years, and fifth grade in 1999). We correlated each school's performance on the SAT-9 with their performance on the state assessment by comparing how the school did on each measure in relation to the other LESCP schools in the state.[9] Table 2-8 shows the results of this analysis. We correlated performance on the SAT-9 with performance on a state assessment in all seven LESCP states.

As shown in Table 2-8, there was variation in the degree to which a school's performance on the SAT-9 was correlated with its performance on the state assessment. This held true both across the states in the LESCP sample and, to some extent, across grade levels.

---

[9] For each school, in both reading and mathematics, we computed the z-score, which placed the school within a distribution of other LESCP schools in their state for each SAT-9 and state assessment score. We then correlated the SAT-9 and state assessment z-scores with each other to determine which tests were correlated with each other.

Table 2-8. Significant correlation coefficients among school rankings within the LESCP sample on the SAT-9 and on state assessment scores

	Reading	Mathematics
State A		
Third grade (1997)	**.71**	
Fourth grade (1997)		**.79**
Fourth grade (1998)		**.83**
Fourth grade (1999)		**.72**
State B		
Third grade (1997)	**.80**	.32
Fifth grade (1999)	**.78**	**.79**
State C		
Fourth grade (1997)	.18	
Fourth grade (1998)	.49	
State D		
Fourth grade (1998)	.63	
Fourth grade (1999)	**.84**	
Fifth grade (1999)		Sample too small for analysis
State E		
Fifth grade (1999)	**.91**	.70
State F		
Third grade (1997)	.58	**.69**
Fifth grade (1999)	**.91**	**.82**
State G		
Third grade (1997)	**.79**	-.32
Fourth grade (1997)	**.96**	.72
Fourth grade (1998)	-.72	-.49
Fourth grade (1999)	.68	.39
Fifth grade (1999)	.73	.48

Table reads: In State A, the state assessment was administered to students at the third-grade level in reading in 1997 and at the fourth-grade level in mathematics in 1997, 1998, and 1999. We found a positive correlation of .71 in the school's ranking within the LESCP sample on its state assessment and SAT-9 results for third-grade reading. This relationship was significant at the .05 level.

Note: Bold type indicates that correlation coefficients were significant at the .05 level. Darkened cells indicate that the state assessment was not administered at the specified grade level in that year.

In five states (all except States C and G) there was a good correlation between relative rankings by SAT-9 and relative rankings by the state assessments. This gives us some confidence that SAT-9 measures achievement in much the same way as many of the state assessments. It suggests that the SAT-9 may provide a good substitute for state assessments in the analysis of student achievement, although it clearly does not map perfectly onto the content and skills measured by all states. This study's use of a test that was uniform across all participating schools, which was necessary to conduct the planned analyses, thus seems to have been a reasonable choice.

Conclusions

The SAT-9 tests offered information about several aspects of student performance in reading and mathematics. They showed, for example, that the LESCP sample as a whole performed below national and urban norms on these tests. These test data also revealed a persistent, negative relationship between poverty—both individual and school-level—and student achievement. Although the SAT-9 measured somewhat different skills than any particular state test, the standardized tests did offer a comparable basis for measuring performance and growth across all the study's classrooms. Thus, they provided much of the data for the analyses described in Chapter 3, which focuses on those instructional conditions that might potentially alleviate the harmful effects of poverty on student performance.

CLASSROOM AND SCHOOL VARIABLES RELATED TO STUDENT PERFORMANCE

We now examine the relationship between classroom and school practices (see Box 3 in Figure 1-1) and student achievement (see Box 4 in Figure 1-1). The chapter addresses the question: To what extent do particular instructional practices or instructional supports show statistically significant relationships with the Longitudinal Evaluation of School Change and Performance (LESCP) students' outcomes in reading and mathematics? The analyses presented in this chapter relate to the achievement of the LESCP longitudinal students in the closed-ended reading and mathematics assessments. We begin by describing the variables used in our analyses, including the characteristics of instruction and instructional support measured by our teacher surveys. We then describe our procedures for tracing the relationship between these variables and students' performance on tests across the 3 years of the study, with variables like student and school poverty used as controls. Because much of this chapter relies on hierarchical linear modeling (HLM) analyses, we outline the key assumptions and procedures associated with that analytic method. Next, we present our models and findings for reading and mathematics.

Variables and Methods Used in the Analysis of Student Performance

Our models assume that both proximal variables and control variables may be directly related to student performance. *Proximal variables* have to do with instruction and are potentially changeable by reform initiatives. They include measures of teacher beliefs, classroom practices in curriculum and instruction, and instructional supports like professional development. We built the proximal variables from the survey responses of individual teachers as well as average values for groups of teachers, either the whole-grade level or the whole school. The reason for considering school-level variables as proximal variables in our statistical models is that the average responses found between groups of teachers, such as all kindergarten to fifth-grade (K–5) teachers in the school, tell us about the overall kind of academic press experienced by students in that school. We hypothesize that school environment as it relates to these proximal variables has an independent effect on student achievement above and beyond individual teacher practices. In short, whether found at the level of the classroom or the surrounding organization, the proximal variables are instructional conditions and supports that may most directly affect student learning. They are the targets of Title I policy and of standards-based reform, and, in Chapter 4, we will report on the extent to which reform policies are affecting them.

In the sections dealing with reading and mathematics, we present detailed descriptions of the proximal variables for each subject area, built from teachers' survey responses. A quick overview, however, is that these variables pertained to specific parts of an overall vision of standards-based reform: a framework of content and performance standards, together with assessments and curriculum keyed to those standards, that would command attention and guide classroom practice; curriculum and instruction designed to engage students in relatively advanced academic tasks rather than bogging down in rote drill and practice; teachers prepared to teach in new ways, having participated in professional development geared to the standards and assessments; and active communication between school and home.[1]

Control variables are also characteristics of the student or school, but they are not specifically related to instruction and are not amenable to intervention by Title I policy or standards-based reform. They include student- and school-level measures of poverty and initial achievement, and school size. Our control variables include the following:

- **Student Poverty** – whether an individual student was eligible to receive free or reduced-price lunch in 1998;
- **School Poverty** – the percentage of longitudinal students in the school eligible to receive free or reduced-price lunch in 1998;
- **Student's Initial Achievement Status** – whether an individual student's third-grade score on the closed-ended reading or mathematics portion of the Stanford Achievement Test, Ninth Edition (SAT-9) fell in the bottom quarter of the national distribution of scores[2];
- **School's Initial Achievement** – the percentage of longitudinal students in the school whose third-grade scores fell in the bottom quarter on this test; and

---

[1] The interrelated parts of a standards-based reform vision are described in J. O'Day and M.S. Smith. (1993). Systemic reform and educational opportunity, in S. Fuhrman (Ed.) *Designing coherent education policy.* San Francisco: Jossey-Bass, pp. 250-312, and in L.B. Resnick and K.J. Nolan. (1995). Standards for education, in D. Ravitch (Ed.), *Debating the future of American education: Do we need national standards and assessments?* Washington, D.C.: Brookings Institution. Other sources for the variables measured in this study included work on curriculum and instruction, such as A. Porter. (1998). *A curriculum out of balance: The case of elementary school mathematics.* East Lansing, Mich.: Institute for Research on Teaching, College of Education, Michigan State University; and M.S. Knapp and P.M. Shields, (Eds.). (1990). *Better schooling for the children of poverty: Alternatives to conventional wisdom, Volume II: Commissioned papers and literature review.* Washington, D.C.: U.S. Department of Education; and work on parent and family involvement such as S.L. Dauber and J.L. Epstein. (1993). Parent attitudes and practices of involvement in inner-city elementary and middle schools, in A.T. Henderson and N. Berla (Eds.). (1994). *A new generation of evidence: The family is crucial to student achievement*, Washington, D.C.: Center for Law and Education; and J. Ballen and O. Moles. (1994). *Strong families, strong schools: Building community partnerships for learning.* Washington, D.C.: U.S. Department of Education.

[2] This status was used as an indicator of initial achievement instead of the actual third-grade score because the third-grade score is already included in the model as a repeated measure.

BEST COPY AVAILABLE

- **School Size** – the total number of students enrolled in the school in 1997 or the average number of students enrolled in the school between 1997 and 1999.

This study relied most heavily on use of a statistical technique called HLM.[3] For a study like this one, where a key objective is to examine student achievement growth over time, HLM offers several advantages over other modeling techniques. Like multiple regression, HLM builds a model in which the effect of each independent variable on an outcome variable can be isolated and measured, by holding constant all the other independent variables. However, HLM addresses the problems of aggregation bias, misestimated standard errors, and heterogeneity of regression. For example, HLM can handle data from multiple levels. This is important because (1) we have a sample of students tracked over time who are nested within a sample of schools and (2) we are interested in assessing the independent effects of student- and school-level factors on achievement. HLM also accounts for the statistical dependence between the outcomes of students who attend the same schools and share educational experiences, and HLM calculates the correct standard errors accordingly. In addition, given that differences between schools might influence the relationship between student characteristics and achievement, HLM estimates separate models of this relationship for each school. Finally, unlike other conventional multivariate, repeated-measure methods, HLM can explicitly model individual growth over time and is more flexible regarding data requirements for this model such as the number of observations that can vary per person.[4] Thus, we were able to simultaneously predict students' third-grade test scores and the rate of increase in their scores from third grade to fifth grade; we were also able to determine the effects of student- and school-level variables on both initial achievement and growth over time.

HLM was used in the LESCP to estimate a trajectory of student outcomes—student test scores in the third through fifth grade. To make an as accurate estimate as possible, we tested all of the variables we thought would be relevant to this trajectory, including third-, fourth-, and fifth-grade teaching practices.

We note that the growth modeling approach we took was consistent with that taken by researchers before us. For example, Prospects used HLM in a similar way and reported on results of teaching practices and other factors on initial score and learning rate.[5] Bryk and Raudenbush analyze and report on their analyses of student achievement in a like manner.[6]

---

[3] We performed two additional types of analyses to corroborate findings from the HLM: analyses of scores for consecutive fourth grades and analyses of school gains on state tests. These analyses are reported on in Appendix C.

[4] See A.S. Bryk and S.W. Raudenbush. (1992). *Hierarchical linear models: Applications and data analysis networks*. Newbury Park, Calif.: Sage, pp. 133-134 for a complete discussion.

While third-grade test scores are a "baseline measurement" in LESCP, in HLM these scores are merely one of the points in the growth trajectory. It is perhaps useful to distinguish between the predicted value (often called latent value) of third-grade scores from the actual third-grade test scores achieved by the students. Both are estimates of the students' actual knowledge. Any test score is an imprecise measure of student achievement. Scores predicted by HLM are probably better measures of achievement because the model reduces testing error by taking into account multiple tests (repeated measures) by the same student as well as other student information.

We did several tasks in the HLM analysis of performance in each subject. We built indices of instruction and instructional support in that subject for individual teachers. We then grouped teachers by all teachers in a school and by all third- through fifth-grade teachers in a school. We then developed a series of growth models for student performance as a function of the indices and control variables. By looking at the results of successive models, we learned which variables mattered and how much they mattered. The first model was "unconditional" with no proximal or control variables; it showed what portion of all the variation in student outcomes occurred between or within schools. Finding substantial variation in student outcomes, we then proceeded to search for the reasons. The second model began that search. It was conditional on the control variables, such as school and student poverty. It showed whether these variables affected the initial scores or the rate of growth, and how much of the variation in scores the control variables explained. In both reading and mathematics, this conditional model did reveal a significant relationship between poverty and initial achievement, but significant variation across students remained, even after controlling for student and school-level poverty.

We built many models that sought to explain the remaining variation across student outcomes by adding in the proximal variables. A series of these models provided the preliminary results that enabled us to build the final model. The more preliminary series of models took each proximal variable one at a time, 1 year of data at a time, measured for individual students (i.e., the student's teacher in each year) and the school (all teachers of longitudinal students and also all teachers in grades K–5).

---

[5] See K.K., Wong, L.V., Ledges, G.D., Borman, and J.V., D'Agostino. (1996). *Prospects: Special analyses. Final report.* Washington, D.C.: U.S. Department of Education, pp. 37-85.

[6] Some of the work is summarized in A.S., Bryk, and S.W., Raudenbush. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, Calif.: Sage Publications, pp. 185-196.

A further refinement on these models was the addition of interaction terms—school poverty with each teaching practice and student's initial achievement status with each teaching practice—to find out whether the instructional practices differentially affected students whose schools had different poverty rates or who had varying levels of initial achievement. We used the Schwarz Bayesian Criterion (BIC) to determine which model had the "best fit" for each teaching practice. Teaching practices that were significant in the best fitting model were retained for the final model. In addition, because of the strong theoretical arguments for the interactions tested, any interaction that was not in the best fitting model for a given teaching practice but reached a moderate level of significance (0.01 or better) was retained for the final model. The contents and results of those final models are reported in this chapter. Details of the approach that was used appear in Appendix A.

The HLM analyses focus on a subset of the LESCP population: the longitudinal students, for whom we have test data from the third, fourth, and fifth grades. As discussed in Chapter 2, those students who did not change schools during the study period scored higher on the SAT-9 than their more mobile classmates. HLM allows the user to keep students in the model for whom we do not have test scores for all 3 years. Including these students in the analyses may make sense if the students were in the school for 3 years but just did not take the test for 1 of the years. However, in LESCP the vast majority of students for whom we do not have three measurement points had left their schools. At the points in time where we do not have test scores, we have no measures for these students for either the dependent variables (test scores) or the independent variables (teaching practices). Thus, we would have had to impute teaching practices for these students even though we had no idea what school they were in. We decided this was not a good idea and decided not to do so. However, while not representative of the full LESCP sample (let alone of all students in high-poverty schools nationally), the longitudinal students provide a valuable source of evidence about the impact of particular instructional practices or supports on student performance because it is the students who remain in a school longest who have the best chance of being affected by that school.

For this report, we also focused the analyses on the closed-ended reading scores for both reading and mathematics. In the course of conducting the study, we obtained both open-ended and closed-ended achievement scores from the SAT-9. The open-ended scores are normed on a smaller sample of students than are the closed-ended scores. The open-ended scores are also more subjective than the closed-ended scores because they are based on scorer ratings of the students' responses to the test items. For both of these reasons, we had less confidence in the ability of the open-ended scores, as compared with the closed-ended scores, to accurately reflect student achievement. However, we will be reporting on our analyses of the relationships among teaching practices and open-ended scores in a separate volume.

HLM Analysis and Results in Reading

The HLM analyses of student performance in reading used individual students' initial scores and gains on the SAT-9, closed-ended, reading test as the outcomes to be explained. The first model in the HLM analysis, the unconditional model, told us where the variation existed in student outcomes. It showed, first of all, that third-grade reading scores varied more than the rate of increase in scores. About three-quarters of this variation in third-grade scores was found within schools (i.e., there were differences across students within schools). One-quarter of the variation was between schools. The model also showed a very different pattern in the variation that did exist in the rate of score gains: virtually all of that variation was between schools (i.e., the variation in the rate of score gains within schools was not statistically significantly different than zero).

Adding the control variables as predictors did not reduce much the largest component of the variance, which was found in the third-grade reading scores **within** schools; the reduction in this variance was just 2.2 percent. These variables did reduce the variance in third-grade reading scores **between** schools by 38.8 percent. They reduced the variance in score increases between schools by 4 percent. This can be understood in the context of the Chapter 2 findings of a strong relationship between poverty and initial score but not between poverty and score gains.

The final model to be described below, which included selected proximal variables along with the control variables, succeeded in further reducing the variance of outcomes, especially between schools. This model still explained only 4.2 percent of the variance in third-grade reading scores within schools. However, it explained 56.4 percent of the variance in third-grade scores between schools, and it reduced the variance in the rate of increase between schools by nearly 40 percent.

The final model is the one whose results we present in detail here. In the preliminary steps taken toward building the final model, where we tried out a large number of proximal variables, we used the following indices built from teachers' survey responses (along with the control variables of poverty, school size, and initial achievement) as possible explanations for initial scores and gains[7]:

- **Visibility of standards and assessments,** an index that gathers together teachers' responses regarding their awareness and use of content standards, performance standards, curriculum frameworks, and required assessments in reading (see Box 3-1 for items included and descriptive statistics).
- **Basic instruction in upper grades,** measuring the extent to which teachers in third through fifth grades had their students engage in instruction at a fairly rudimentary level (Box 3-2). The 0-135 scale represents an estimate of the number of lessons per year for each activity listed.

Box 3-1. Visibility of standards and assessments (reading)

Items:

Please indicate how familiar you are with each of the following for your state or district:

1. **Content standards in reading/language arts (scale: 1–4)**
2. **Curriculum frameworks in reading/language arts (scale: 1–4)**
3. **Student assessments in reading/language arts (scale: 1–4)**
4. **Performance standards in reading/language arts (scale: 1–4)**

1 = Not at all familiar; 2 = A little familiar; 3 = Moderately familiar; 4 = Very familiar

To what extent do you believe the reading/language arts curriculum you teach reflects each of the following?

5. **Content standards (scale: 1–4)**
6. **Curriculum frameworks (scale: 1–4)**
7. **Student assessments (scale: 1–4)**
8. **Performance standards (scale: 1–4)**

1 = Not at all; 2 = Small extent; 3 = Moderate extent; 4 = Great extent

Descriptive statistics:

Maximum possible score: 32 Minimum possible score: 8

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	26.8	4.0	32.0	21.0	78%	22%
1998	27.2	3.9	32.0	23.0	76%	24%
1999	27.8	3.6	32.0	24.0	67%	33%

---

[7] See Appendix D for reliability measures for these indices.

Box 3-2. Basic instruction in upper grades (reading)

Items:

For each of the reading/language arts student lesson activities listed below, indicate how frequently you have your students engage in it, and if you do use the activity, how much per lesson:

1. Read aloud (scale: 0–135)
2. Complete reading workbooks or skill-sheet assignments (scale: 0–135)
3. Practice phonics (scale: 0–135)
4. Practice word attack (scale: 0–135)

Scale is an estimate of the number of lessons per year.

Descriptive statistics:

Maximum possible score: 540 Minimum possible score: 0

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	171.4	112.3	324.0	48.4	75%	25%
1998	185.2	120.8	360.0	54.0	76%	24%
1999	182.6	125.1	360.0	47.3	74.9%	25.1%

- **Preparation for instruction,** or the extent to which teachers felt that they were well prepared to use particular strategies in reading instruction (Box 3-3).

- **Rating of professional development,** combining teachers' ratings of the quality of reading-related professional development that they had received over the past 12 months (Box 3-4); this index could have a value of zero when the teacher or all teachers at a grade level had had no professional development in reading over that period, but at the K–5 school level it is built only from the responses of those teachers who had participated in professional development.

- **Outreach to low achievers' parents,** indicating how many of the parents of struggling students the teacher had tried to reach in various ways (Box 3-5).

Box 3-3. Preparation for instruction (reading)

Items:

How well prepared are you to do each of the following?

1. Use small group instruction for reading/language arts (scale: 1–4)
2. Take into account students' existing skills levels when planning curriculum and instruction (scale: 1–4)
3. Integrate reading/language arts into other content areas (scale: 1–4)
4. Use a variety of assessment strategies (scale: 1–4)
5. Teach groups that are heterogeneous in ability (scale: 1–4)

1 = Not well prepared; 2 = Somewhat prepared; 3 = Fairly well prepared; 4 = Very well prepared

Descriptive statistics:

Maximum possible score: 20 Minimum possible score: 5

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	17.9	2.4	20.0	15.0	68%	32%
1998	18.0	2.4	20.0	15.0	66%	34%
1999	18.0	2.5	20.0	15.0	65%	35%

Box 3-4. Rating of professional development (reading)

Items:

What was the quality of professional development you received in the past 12 months on:

1. Content in reading (scale: 0–3)
2. Instructional strategies for teaching reading (scale: 0–3)

0 = Did not have; 1 = Low quality; 2 = Adequate quality; 3 = High quality

To what extent was the professional development activity:

3. Well matched to your school's or department's plan to change practice (scale: 0–5)
4. Designed to support reform efforts underway in your school (scale: 0–5)
5. Designed to support state or district standards or curriculum frameworks (scale: 0–5)
6. Designed to support state or district assessment (scale: 0–5)

0 = Did not have; 1–5 scale from "not at all" to "great extent"

To what extent were your knowledge and skills enhanced in each of the following ways as a result of your participation in the professional development experiences you have had in the past year:

7. Helped me adapt my teaching to meet state assessment requirements (scale: 0–5)
8. Helped me adapt my teaching to meet state standards or curriculum framework requirements (scale: 0–5)
9. Gained confidence in using new pedagogical approaches in teaching reading/English language arts (scale: 0–5)

0 = Did not have; 1–5 scale from "not at all" to "great extent"

Descriptive statistics:

Maximum possible score: 41 Minimum possible score: 0

	Mean	Standard Deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1998	23.9	14.9	40.0	0.0	89%	11%
1999	25.8	13.3	39.0	0.0	89%	11%

Box 3-5. Outreach to low achievers' parents (reading)

Items:

For how many of your low-achieving students did you do each of the following?

1. Initiate face-to-face meetings with parents (scale: 1–4)
2. Initiate telephone calls to parents when their child was having problems (scale: 1–4)
3. Initiate telephone calls to parents when their child was not having problems (scale: 1–4)
4. Send materials to parents on ways they can help their child at home (scale: 1–4)

1 = Few or none; 2 = Some; 3 = Many; 4 = Most

Descriptive statistics:

Maximum possible score: 16 Minimum possible score: 4

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	11.5	2.9	16.0	8.0	74%	26%
1998	11.6	2.9	16.0	8.0	71%	29%
1999	11.5	2.9	16.0	8.0	76%	24%

Through testing the control variables and the above indices for their relationship with student achievement in a series of relatively simpler models, we selected the combination of significant variables, measured at different levels and in different years, for inclusion in the final reading model.[8] A final combined HLM model simultaneously tested the effects of this combination of variables on students' initial, third-grade score and on their score gains from third to fifth grade. We reiterate that the HLM results show statistically significant relationships among the proximal variables and third-grade score and score gains from the third to the fifth grade. The analyses do not prove causality (i.e., a teaching practice causes an increase in test score) because we do not know if there were other, intervening factors or if students with higher or lower scores were taught differently because of their differing initial levels of achievement. (A random assignment design would more likely yield stronger statements of causality, but this was not the design of LESCP.) We stress this caveat to avoid misinterpretation of the findings. The statistically significant relationships are important findings and allow us to conjecture about causality, and we have done so in this chapter where it is appropriate. We especially avoid assuming causality when we discuss the relationships between third-grade teaching practices and third-grade scores. This is because without a pre-third-grade baseline, we are less sure of whether students of different initial achievement were systematically placed in classrooms with different curricula or instructional practices.

For simplicity of presentation, Table 3-1 displays the measures used to predict students' gains on closed-ended reading test scores from the third grade to the fifth grade, and Table 3-2 displays the measures used to predict students' third-grade score.

Using all of the variables listed in Table 3-1, we predicted longitudinal students' growth on the SAT-9, closed-ended, reading test between third and fifth grade. Not all of the variables were related in a statistically significant way with growth in achievement. For those variables that were significant, Table 3-3 shows the difference in points gained between third and fifth grade for students whose teachers or schools were at the top or bottom of the indices that proved to be significant, holding all other effects at their average.

---

[8] See Appendix A (Sections A.1.3 and A.1.4) for a more complete discussion of the process of including and excluding variables from the final model.

Table 3-1. Variables used to predict each longitudinal student's learning rate in the final HLM reading model

---

Control variables

- Student poverty
- School poverty
- Average school enrollment
- Student's initial achievement status
- School's initial achievement

Student-level variables

- Basic instruction in upper grades
  - Fifth-grade teacher (1999)

- Outreach to low achievers' parents
  - Third-grade teacher (1997)

School-Level Variables

- Visibility of standards and assessments
  - Whole school (1998)

- Basic instruction in upper grades
  - All fifth-grade teachers (1999)

- Rating of professional development
  - Whole school (1999)

- Outreach to low achievers' parents
  - All third-grade teachers (1997)

To give a sense of how much difference in score gains was associated with each variable, we show the number of test-score points above or below the third-grade average for a hypothetical student whose classroom or school was at the 90th or the 10th percentile for the LESCP sample on that particular variable and whose classroom or school was at the average value for every other variable. We selected the 90th- and 10th-percentile values for presentation purposes only. The underlying relationship between the variable and achievement growth remains the same regardless of the percentile value selected. Larger numbers in Table 3-3 mean that a very high or low value on that variable, relative to the whole LESCP sample, was associated with more of a difference in students' achievement growth, holding all other variables at their average.

Table 3-2. Variables used to predict each longitudinal student's third-grade score in the final HLM reading model

---

Control variables

- Student poverty
- School poverty
- School size in 1997

Student-level variables

- Visibility of standards and assessments
    - Third-grade teacher (1997)

- Outreach to low achievers' parents
    - Third-grade teacher (1997)
    - Third-grade teacher (1997), interacting with school poverty

School-level variables

- Visibility of standards and assessments
    - All third-grade teachers (1997)
    - Whole school (1997)
    - Whole school (1997), interacting with school poverty

- Preparation for instruction
    - Whole school (1997)
    - Whole school (1997), interacting with school poverty

- Outreach to low achievers' parents
    - All third-grade teachers (1997)
    - All third-grade teachers (1997), interacting with school poverty

---

Findings for Growth in Reading Scores

- **Control variables of student poverty and the initial test score in the final model were not related to students' gains on the reading tests between third and fifth grades. Poverty had no effect on the rate of growth in scores, whether poverty was measured for the individual student or for the school as a whole.** Similarly, students with scores falling nationally in the bottom quarter on the third-grade test had increases that paralleled those of other students.

- **Student-level classroom variables associated with different rates of gain were basic instruction in upper grades and outreach to low achievers' parents. Students whose fifth-grade teachers spent relatively more time offering reading instruction at a basic level gained slightly less, on average, than those whose teachers spent less time working at that level.** For example, students whose fifth-grade teacher was at the 90th percentile of our index of basic instruction gained 1.9 fewer points over 2 years than students whose teachers had an average emphasis on this technique. Students whose fifth-grade teacher was at the 10th percentile of this index gained 1.8 points more over 2 years than the average longitudinal student, holding all other effects at their average. Thus, growth in test scores was 10 percent lower (36.4 versus 40.1 over the 2 years) when teachers spent a lot of time on basic instruction than when they spent little time engaged in those activities. The direction of this relationship is unclear. Possibly, excessive use of basic instruction in the upper grades retards growth. Because we controlled for initial (third grade) score in the model and also examined the interaction of third-grade score with this variable, we believe this explanation is highly plausible. This interpretation is consistent with research and advocacy of the past 10 years.[9] Alternatively, teachers whose students are poor achievers feel the need to reinforce basic skills.

Table 3-3. Final reading HLM model: Effects on score gains of longitudinal students, significant independent variables only

Average gains		
Points gained, third to fifth grade, under average conditions on all independent variables	38.3[a]	
Difference in 2-year gain associated with high/low value on each significant independent variable		
Student-level variables	**Teacher at the 90th percentile**	**Teacher at the 10th percentile**
Basic instruction in upper grades (fifth-grade teacher, 1999)	-1.9[c]	+1.8[c]
Outreach to low achievers' parents (third-grade teacher, 1997)	+4.6[a]	-4.3[a]
School-level variables	**School at the 90th percentile**	**School at the 10th percentile**
Rating of professional development (whole school, 1999)	+3.1[c]	-4.0[c]
Outreach to low achievers' parents (all third-grade teachers, 1997)	+3.7[b]	-3.7[b]

Table reads: For average conditions on all independent variables, the model predicts a 2-year gain of 38.2 points. Students whose fifth-grade teacher was at the 90th percentile in the use of basic instruction in upper grades would gain 1.9 points less; students whose fifth-grade teacher was at the 10th percentile on this variable would gain 1.8 points more.

[a] Significant effect at the .001 level

[b] Significant effect at the .01 level

[c] Significant effect at the .05 level

---

[9] See, for example, Education Trust. (1999). *Dispelling the myth: High poverty schools exceeding expectations.* Washington, D.C.: Author; and Knapp, Shields, and Turnbull, op. cit., footnote 1.

Third-grade reading teacher's outreach to parents of low-achieving students, although associated with lower third-grade performance on the SAT-9, was also associated with more pronounced growth between third and fifth grade. Later in this section, we will see that more outreach by third-grade reading teachers was associated with lower third-grade reading scores. We interpret this association to mean that teachers recognized the need for outreach when there were many low achievers in their classes. Having done so, the outreach to these students' parents appears to have a significant beneficial effect on growth in performance after the third grade. By fifth grade, students whose third-grade teacher contacted more parents of low-achieving students gained back 4.6 points of the 5.1-point deficit that they started out with at the end of the third grade.

- **School variables associated with different rates of gain were the rating of professional development and outreach to low achievers' parents. Students in schools where teachers overall gave relatively favorable ratings to their professional development in reading from spring 1998 through spring 1999 gained more points on the SAT-9 than students in schools where the teachers gave average ratings to that professional development.** The survey questions on professional development assessed both its quality and its fit with standards and assessments, thus yielding some support for the argument that professional development should be geared toward those policies. (See articles by Smith, 1993 and by Resnick and Nolan, 1995, which are cited in footnote 1.) In schools at the 90th percentile on this measure, students' average 2-year gains were 3.1 points more than average; in schools at the 10th percentile they were 4 points less. This translates into a 20 percent greater growth in test scores between grades three and five at schools where teachers rated their professional development high than when they gave it a low rating.

  Schools with relatively high levels of outreach to the parents of low-achieving students, measured for all the teachers in the third grade in 1997, also had students with higher rates of growth. When all third-grade teachers in a school contacted more of the parents of low-achieving students, students gained 3.7 more points on the SAT-9 by the fifth grade than their counterparts in schools where third-grade teachers had average levels of outreach, holding all other effects at their average. Therefore, students benefited when they experienced a combination of (1) a third-grade teacher who emphasized parent outreach for low-achieving students and (2) an emphasis on parent outreach across all third-grade teachers in the school. Growth in test scores between third and fifth grade was 50 percent higher for those students whose teachers and schools reported high levels of parental outreach early than students whose teachers and schools reported low levels of parent outreach activities for the third grade.

Next we discuss the associations among the variables listed in Table 3-2 and longitudinal students' third-grade test scores in reading. Table 3-4 shows the difference in third-grade scores for students whose teachers or schools were at the 10th percentile or 90th percentile of the LESCP sample as measured by our indices, holding all other variables at their average. It shows only the effects of those variables that had significant effects in the final model, dropping a number of variables that were shown in Table 3-2 but that proved not to be significantly related to student achievement in this final multivariate analysis.

Table 3-4. Final reading HLM model: Effects on third-grade achievement of longitudinal students, significant independent variables only

Average third-grade score				
Third-grade score of a student experiencing average conditions on all independent variables	608.6[a]			
Difference in score associated with high/low value on each significant independent variable				
Control variables	**Student eligible for free or reduced-price lunch or school at the 90th percentile**		**Student not eligible for free or reduced-price lunch or school at the 10th percentile**	
Student poverty	-6.1[a]		+8.5[a]	
School poverty	-11.8[a]		+11.6[a]	
Student-level variables	**Teacher at the 90th percentile**		**Teacher at the 10th percentile**	
Visibility of standards and assessments (third-grade teacher)	+2.8[b]		-2.9[b]	
Outreach to low achievers' parents (third-grade teacher)	-5.1[a]		+4.8[a]	
Outreach to low achievers' parents (third-grade teacher) interacting with school poverty	**And 90th percentile school poverty**	**And 10th percentile school poverty**	**And 90th percentile school poverty**	**And 10th percentile school poverty**
	-8.5[c]	-1.5[c]	+7.9[c]	+1.4[c]
School-level variables	**School at the 90th percentile**		**School at the 10th percentile**	
Visibility of standards and assessments in 1997 (whole school) interacting with school poverty	**And 90th percentile school poverty**	**And 10th percentile school poverty**	**And 90th percentile school poverty**	**And 10th percentile school poverty**
	+9.1[b]	-9.8[b]	-10.7[b]	+11.6[b]

Table reads: For longitudinal students with average values on all independent variables, the model predicts a third-grade score of 608.6 on the SAT-9, closed-ended reading test. Students eligible for free or reduced-price lunch would score 6.1 points less; students not eligible for free or reduced-price lunch would score 8.5 points more.

[a] Significant effect at the .001 level

[b] Significant effect at the .01 level

[c] Significant effect at the .05 level

BEST COPY AVAILABLE

Findings for Third-Grade Reading Scores

- **Control variables of student- and school-level poverty had strong negative associations with student outcomes in third-grade reading[10]:**
  - Students eligible for free or reduced-price lunch scored 6.1 points lower than the score predicted by this model for all longitudinal students in the sample. In contrast, their nonpoor classmates scored 8.5 points above the average score.
  - School poverty had an independent negative effect on third-grade achievement. In schools at the 90th percentile of school poverty in our sample, the students scored 11.8 points below average; students in schools at the 10th percentile on this measure scored 11.6 points above average.

- **Student-level proximal variables related to third-grade scores were the visibility of standards and assessments and outreach to low achievers' parents. Third-grade scores tended to be higher in classrooms with higher visibility of standards and assessments.** These were classrooms where teachers were aware of and implementing aspects of standards-based reform—content standards, curriculum frameworks, performance standards, and student assessments. For third-grade teachers at the 90th percentile on this index, relative to the average value for this sample, the model indicated that students would score 2.8 points above the overall average; students whose teachers were at the 10th percentile on this measure would score 2.9 points below average.

  The model shows a negative correlation between student achievement in the third grade and teachers' reported outreach to the parents of low-achieving students. That is, the lower the third-grade score the more the teacher tried to involve low-achieving students' parents. We do not interpret this to mean that more effort by a teacher to involve parents of low achievers lowers a test score. Rather, we interpret this to mean that teachers correctly identified the low achievers in their classes and exerted extra effort in involving the parents of these children. Holding other variables at their average—including school poverty—if a student's third-grade teacher was at the 90th percentile of our index of outreach to low achievers' parents, the student would be expected to have scored 5.1 points lower than if his or her teacher had a LESCP average value for this index. Students whose teachers were at the 10th percentile of this index tended to score 4.8 points above average.

  This relationship, however, was more pronounced at higher poverty schools. The model identified a significant and negative interaction between higher values on this index and school poverty as a predictor of third-grade score in reading. Consider again the above example of the student having an expected score 5.1 points lower if his or her teacher was at the 90th percentile on this index compared with a teacher at the average for this index. If this student were also at a school at the 90th percentile of LESCP schools on poverty (100% of longitudinal students eligible for free or reduced-price lunch), the expected difference for a student with a 90th-percentile teacher versus an average teacher on this index increases the deficit by 8.5 points. Compounding the interaction effect is the original 11.8-point deficit associated with being in a high-poverty school, which leaves such a student below the sample average by 25.4 points.

---

[10] As noted in Chapter 2, the measure of poverty we used was eligibility for free or reduced price lunch. Other, more precise, measures of poverty, had they been available, may have yielded different results.

- **At school level, the visibility of standards and assessments among all K–5 teachers in the school was significantly associated with students' third-grade scores, although this effect was found only in interaction with school poverty.** Overall, students in higher poverty schools, score 11.8 points less than the average for third-grade reading. However, if these students attend schools with higher visibility of standards and assessments, this difference is reduced by 9.1 points.

All of these findings are framed in terms of relatively high or low values of each variable, relative to the LESCP sample. Thus, we briefly describe the overall distribution of teachers' responses in 1999, with attention to the specific response patterns found at the high and low ends for each variable that made a difference in student performance.

Overall, teachers reported a generally high level of familiarity with and integration of reading and language arts standards into their curriculum. The mean score on our Visibility of Standards and Assessments Index was 27.8 in 1999, on the high end of the possible range of 8 to 32. Teachers at the 90th percentile of the index, who we reported to have high visibility of standards and assessments, were at the maximum of 32, answering every question at the high end of the scale, indicating that they were very familiar with standards and assessments and felt that they were strongly reflected in the classroom curriculum. In comparison, teachers at the 10th percentile had an index score of 24, still well above the lowest possible of eight.

The mean score of 11.5 (on a scale of 4 to 16) on our Outreach to Low Achievers' Parents Index suggests that the average teacher in our sample reported initiating contact with parents of **many** of their low-achieving students, through face-to-face meetings, telephone calls, or by sending home materials. Teachers who fell at the 90th percentile had the maximum index score of 16, meaning that they reported initiating communication with the parents of **most** of their low-achieving students. In contrast, the score of eight for teachers at the 10th percentile indicates contact with fewer, but still **some**, parents of low-achieving students.

In general, teachers reported a moderately low frequency and/or intensity of use of activities like reading aloud, completing reading workbooks, or practicing phonics or word attack, reflected in the mean score of 182.6 (on a scale of 0 to 540 lessons per year) for the Basic Instruction in the Upper Grades

Index. Teachers who we considered to emphasize basic instruction in the upper grades reported use of these activities somewhat more frequently or intensively; the index score at the 90th percentile was 360 lessons per year, well below the maximum score of 540, but still considerably higher than teachers at the 10th percentile who scored only 47.3.

Teachers' ratings of the quality of reading professional development they received, and of the extent to which it was consistent with their districts' standards and assessments, displayed wide variation. The average score on our Rating of Professional Development in Reading Index was 25.8 on a scale of 0 to 41. We considered teachers at the 90th percentile to have a high rating of professional development (index score of 39), suggesting a high level of satisfaction with the quality of reading professional development and with the extent to which it matched reform efforts and enhanced their knowledge and skills. In contrast, teachers at the 10th percentile fell at the lowest possible index score of zero, indicating that they had received no professional development in reading over the past year.

We looked for students in our sample who came closest to experiencing the favorable instructional conditions identified by the model. There were 110 students who experienced values above the LESCP average on the variables that had a significant, positive relationship to reading score gains. They were disproportionately poor and so were their schools; 80 percent of them were eligible for free or reduced-price lunch (compared with 67% of other longitudinal students), and 50 percent were in schools where at least 90 percent of the students were poor (compared with 28%). Table 3-5 presents the performance of those students compared with the average longitudinal student and national norms.

Table 3-5. Reading base-year scores and gains for longitudinal students with favorable instructional conditions, average longitudinal students, and national norms students

	Longitudinal students with favorable instructional conditions	Average predicted by model for longitudinal sample	National norms, 1995 (50th percentile students)
Third-grade reading score	598	608.6	616
Two-year gain	44	38.3	38
Fifth-grade reading score	642	646.9	654

Having started out 10.6 points behind their peers and 18 points below national norms, they closed the gap to end up 4.9 points behind their peers and 12 points below national norms. Thus, even above-average values on the significant variables conferred some benefit on students' 2-year growth, although it was not enough to enable them to catch up with the national norms set in 1995.

HLM Analysis and Results in Mathematics

The HLM method was also used to analyze individual students' initial closed-ended scores and their gains over time for the subject of mathematics. Again, the HLM modeling began with an unconditional model that could tell us where the variation existed in student outcomes. Mathematics resembled reading in all respects here. Third-grade mathematics scores varied more than the rate of increase in scores; about three-quarters of the variation in the third-grade scores was found within schools, while one-quarter was between schools. As with reading, essentially all of the variation in the rate of score gains was between schools.

Adding the control variables as predictors only slightly reduced the largest component of the variance, which was found in the third-grade mathematics scores **within** schools. The reduction in this variance was just 5.7 percent. The control variables reduced the variance in third-grade mathematics scores **between** schools by 18.6 percent. They did not reduce the variance in score increases between schools.

As in reading, the final model, which included selected instructional (proximal) variables along with the control variables, succeeded in further reducing the variance of outcomes, especially between schools. This model explained just 5.6 percent of the variance in third-grade mathematics scores within schools. It explained 26.1 percent of the variance in third-grade scores between schools (less than our final model did in reading). It reduced the variance in the rate of increase between schools by 52.7 percent (more than the final model in reading).

The final model in mathematics is the one whose results we present in detail here. The variables tried out in the process of building the final model included the following indices built from teachers' survey responses (along with the control variables of poverty, school size, and initial achievement status) as possible explanations for initial scores and gains[11]:

- **Visibility of standards and assessments,** an index that gathers together teachers' responses regarding their awareness and use of content standards, performance standards, curriculum frameworks, and required assessments in mathematics (see Box 3-6 for items included and descriptive statistics).

---

[11] See Appendix D for reliability measures for these indices.

Box 3-6. Visibility of standards and assessments (mathematics)

Items:

Please indicate how familiar you are with each of the following for your state or district:

1. Content standards in mathematics (scale: 1–4)
2. Curriculum frameworks in mathematics (scale: 1–4)
3. Student assessments in mathematics (scale: 1–4)
4. Performance standards in mathematics (scale: 1-4)

1 = Not at all familiar; 2 = A little familiar; 3 = Moderately familiar; 4 = Very familiar

To what extent do you believe the mathematics curriculum you teach reflects each of the following?

5. Content standards (scale: 1–4)
6. Curriculum frameworks (scale: 1–4)
7. Student assessments (scale: 1–4)
8. Performance standards (scale: 1–4)

1 = Not at all; 2 = Small extent; 3 = Moderate extent; 4 = Great extent

Descriptive statistics:

Maximum possible score: 32 Minimum possible score: 8

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	26.8	4.0	32.0	22.0	68%	32%
1998	27.4	3.9	32.0	23.0	73%	27%
1999	27.7	3.9	32.0	23.0	64%	36%

- **Exploration in instruction,** providing an overall gauge of the time spent (measured by approximate number of lessons per year) in several instructional activities in which students take initiative or work on more complicated assignments; the measure combines teachers' responses about both the frequency and the duration of these activities (Box 3-7).

- **Presentation and practice in instruction,** combining answers about the frequency and duration of teacher-directed instruction or individual student skill practice (Box 3-8).

- **Preparation for instruction,** or the extent to which teachers felt that they were well prepared to use particular strategies in mathematics instruction (Box 3-9).

- **Rating of professional development,** combining teachers' ratings of the quality of mathematics-related professional development that they had received over the past 12 months (Box 3-10); this index could have a value of zero when the teacher or all teachers at a grade level had had no professional development in mathematics over that period, but at the K–5 school level it is built only from the responses of those teachers who had participated in professional development.

- **Outreach to low achievers' parents,** indicating how many of the parents of struggling students the teacher had tried to reach in various ways (Box 3-11, which shows that this index was built from the same survey items in mathematics as in reading and also shows the descriptive statistics for teachers of mathematics).

Box 3-7. Exploration in instruction (mathematics)

Items:

For each of the types of instructional activities listed below, indicate the extent to which you use it and, if you do use it, how much:

1. Use manipulatives to demonstrate a concept (scale: 0–135)
2. Discuss multiple approaches to solving a problem (scale: 0–135)

For each of the student lesson activities, indicate how frequently you have your students engage in the activity and, if you do use it, how much:

3. Work on problems in small groups to find a joint solution (scale: 0–135)
4. Have whole class discuss solutions developed in small groups (scale: 0–135)
5. Have student-led whole group discussions (scale: 0–135)
6. Represent and analyze relationships using tables and graphs (scale: 0–135)
7. Respond to questions or assignments that required writing at least a paragraph (scale: 0–135)
8. Work with manipulatives (scale: 0–135)
9. Work on projects/assignments that take a week or more to finish (scale: 0–135)

Scale is an estimate of the number of lessons per year.

Descriptive statistics:

Maximum possible score: 1,215 Minimum possible score: 0

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	274.6	188.1	499.5	82.1	70%	30%
1998	295.8	181.4	525.4	102.0	59%	41%
1999	286.9	186.4	528.8	97.1	65%	35%

Box 3-8. Presentation and practice in instruction (mathematics)

Items:

For each of the types of instructional activities listed below, indicate the extent to which you use it and, if you do use it, how much:

1. Lecture or present (scale: 0–135)
2. Lead whole group discussions (scale: 0–135)
3. Demonstrate working an exercise at a board (scale: 0–135)
4. Administer a test (scale: 0–135)

For each of the student lesson activities, indicate how frequently you have your students engage in the activity and, if you do use it, how much:

5. Respond orally to questions on subject matter (scale: 0–135)
6. Work individually on written assignments or worksheets in class (scale: 0–135)
7. Practice or drill on computational skills (scale: 0–135)

Scale is an estimate of the number of lessons per year.

Descriptive statistics:

Maximum possible score: 945 Minimum possible score: 0

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	330.8	143.5	522.0	175.5	75%	25%
1998	352.4	141.6	543.4	196.9	69%	31%
1999	362.7	161.7	576.0	182.3	77%	23%

BEST COPY AVAILABLE

Box 3-9. Preparation for instruction (mathematics)

Items:

How well prepared are you to do each of the following?

1. Present the applications of mathematics concepts (scale: 1–4)
2. Use cooperative learning groups in mathematics (scale: 1–4)
3. Take into account students' previous conceptions about mathematics when planning curriculum and instruction (scale: 1–4)
4. Integrate mathematics with other subject areas (scale: 1–4)
5. Manage a class of students who are using manipulatives (scale: 1–4)
6. Use a variety of assessment strategies (scale: 1–4)
7. Use the textbook as a resource rather than as the primary instruction tool (scale: 1–4)
8. Teach groups that are heterogeneous in ability (scale: 1–4)

1 = Not well prepared; 2 = Somewhat prepared; 3 = Fairly well prepared; 4 = Very well prepared

Descriptive statistics:

Maximum possible score: 32 Minimum possible score: 8

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	27.6	4.3	32.0	22.0	67%	33%
1998	27.9	4.3	32.0	23.0	66%	34%
1999	27.8	4.5	32.0	22.0	67%	33%

- **Use of National Council of Teachers of Mathematics (NCTM) standards.** A single question asked teachers about the extent to which their curriculum followed the standards of the NCTM. In response to interim reports from this study, there was considerable interest in the implications of preliminary findings about teachers' use of standards in mathematics. The study team was asked about the classroom use of NCTM standards and any relationship between teachers' use of these standards and student achievement. The team's first step in responding to these questions was to try to develop an index that combined answers to several questions, including the one about use of NCTM standards; however, other classroom practices reported by teachers did not correlate well with this question. Thus, we consider the measure to be weak relative to the other measures that were built on responses to multiple questions around an issue that were highly correlated with each other.

Once again, through testing the control variables and the above indices for their relationship with student achievement in a series of relatively simpler models, we selected the combination of variables, measured at different levels and in different years, for inclusion in the final mathematics model.[12] A combined HLM model for mathematics simultaneously tested the effects of this combination of variables on students' initial, third-grade score and on their score gains from third to fifth grade. For simplicity of presentation, however, Tables 3-6 and 3-7 display the variables used to predict students'

---

[12] Again, see Appendix A (Sections A.1.3 and A.1.4) for a more detailed explanation of the process of including and excluding variables from the final model.

Box 3-10. Rating of professional development (mathematics)

Items:

What was the quality of professional development you received in the past 12 months on:

1. Content in mathematics (scale: 0–3)
2. Instructional strategies for teaching mathematics (scale: 0–3)

0 = Did not have; 1 = Low quality; 2 = Adequate quality; 3 = High quality

To what extent was the professional development activity:

3. Well matched to your school's or department's plan to change practice (scale: 0–5)
4. Designed to support reform efforts underway in your school (scale: 0–5)
5. Designed to support state or district standards or curriculum frameworks (scale: 0–5)
6. Designed to support state or district assessment (scale: 0–5)

0 = Did not have; 1–5 from "not at all" to "great extent"

To what extent were your knowledge and skills enhanced in each of the following ways as a result of your participation in the professional development experiences you have had in the past year:

7. Helped me adapt my teaching to meet state assessment requirements (scale: 0–5)
8. Helped me adapt my teaching to meet state standards or curriculum framework requirements (scale: 0–5)
9. Gained confidence in using new pedagogical approaches in teaching mathematics (scale: 0–5)

0 = Did not have; 1–5 scale from "not at all" to "great extent"

Descriptive statistics:

Maximum possible score: 41 Minimum possible score: 0

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1998	23.0	15.4	40.0	0.0	88%	12%
1999	21.6	15.4	38.0	0.0	82%	18%

gains, showing separately those measured at the student and school levels, and Table 3-8 displays the variables used to predict students' third-grade scores.

We used all the variables listed in Tables 3-6 and 3-7, above, to predict longitudinal students' 2-year gains on the SAT-9, closed-ended, mathematics test. As seen in Table 3-9, a rich and complicated picture emerged, with many variables showing significant relationships with student gains, either alone or in interaction with control variables.

As in reading, we indicate how much difference in scores was associated with each variable by showing the number of test-score points above or below the third-grade average for a hypothetical student whose teacher or school was at the 90th or the 10th percentile for the LESCP sample on that particular variable, and whose teacher or school was at the average value for every other variable.

Box 3-11. Outreach to low achievers' parents (mathematics)

Items:

For how many of your low achieving students did you do each of the following?

5. Initiate face-to-face meetings with parents (scale: 1–4)
6. Initiate telephone calls to parents when their child was having problems (scale: 1–4)
7. Initiate telephone calls to parents when their child was not having problems (scale: 1–4)
8. Send materials to parents on ways they can help their child at home (scale: 1–4)

1 = Few or none; 2 = Some; 3 = Many; 4 = Most

Descriptive statistics:

Maximum possible score: 16 Minimum possible score: 4

	Mean	Standard deviation	90th Percentile	10th Percentile	% Variance btw. schools	% Variance w/in schools
1997	11.5	2.9	16.0	8.0	74%	26%
1998	11.6	2.9	16.0	8.0	72%	28%
1999	11.5	2.9	16.0	8.0	77%	23%

Larger numbers in Table 3-9 mean that a relatively high or low value on that variable was associated with more of a difference in students' achievement growth, all other things at the average. In some cases, where the effect of a variable was different depending on the level of another variable, we show a more complicated display. When a significant interaction was found with poverty, we show separate columns for the instructional variable's effects on student performance in schools serving very high or low proportions of poor students, relative to the LESCP sample average.

Findings for Growth in Mathematics Scores

- **Unlike the findings for reading, there was a significant relationship between the main effect of school poverty and the rate at which students gained in mathematics.** As we will see later in this section, student and school poverty were both related to lower initial, third-grade test scores in both subjects. In reading, however, there was no main effect between poverty and the rate at which students gained over the 2-year study period. This pattern changed in mathematics, where attending a high-poverty school meant that a student tended to narrow the initial gap, holding all other variables constant. By fifth grade, the model showed, students in schools with poverty levels at the 90th percentile would make gains that were 5.8 points greater than average, while students in schools with relatively low poverty (for this sample) would fall 6.6 points short of the average rate of gain.

  Another control variable, the amount of low achievement found in the school in the base year, also had a significant, independent relationship to student gains. In schools at the 90th percentile of this distribution in the LESCP sample (schools with very high proportions of initially low-achieving students), students' 2-year mathematics gains were significantly less than average—10.8 points less, in comparison to a school at the opposite end of the LESCP sample distribution on initial achievement, where the gains would be 9.4 points greater than average.

Table 3-6. Variables used to predict each longitudinal student's learning rate in the final HLM mathematics model, control and student-level variables

---

Control variables

- Student poverty
- School poverty
- Average school enrollment
- Student's initial achievement status
- School's initial achievement

Student-level variables

- Visibility of standards and assessments
  - Fourth-grade teacher (1998)
  - Fourth-grade teacher (1998), interacting with school poverty
  - Fifth-grade teacher (1999)
- Reported use of NCTM standards
  - Third-grade teacher (1997)
  - Third-grade teacher (1997), interacting with school poverty
- Exploration in instruction
  - Fifth-grade teacher (1999)
  - Fifth-grade teacher (1999), interacting with student's initial achievement status
- Presentation and practice in instruction
  - Third-grade teacher (1997)
- Preparation for instruction
  - Fifth-grade teacher (1999)
- Rating of professional development
  - Fifth-grade teacher (1999)

Student-level instructional variables that showed a relationship to the rate of gains for all students were exploration in instruction and the teacher's rating of professional development. In addition, student-level measures of the visibility of standards and assessments and the reported use of NCTM standards had effects only in conjunction with school poverty. Exploration in instruction, as reported by the student's fifth-grade teacher, was positively related to student gains when all other conditions were at their average values. If the teacher reported a great deal of use of

Table 3-7. Variables used to predict each longitudinal student's learning rate in the final HLM mathematics model, school-level variables

---

School-level variables

- Visibility of standards and assessments
  - All fourth-grade teachers (1998)
  - All fourth-grade teachers (1998), interacting with school poverty
  - All fifth-grade teachers (1999)
  - Whole school (1997)

- Reported use of NCTM standards
  - All third-grade teachers (1997)
  - All third-grade teachers (1997), interacting with school poverty

- Exploration in instruction
  - All fifth-grade teachers (1999)
  - All fifth-grade teachers (1999), interacting with student's initial achievement status

- Presentation and practice in instruction
  - All third-grade teachers (1997)
  - All fifth-grade teachers (1999)
  - All fifth-grade teachers (1999), interacting with school poverty

- Preparation for instruction
  - All fifth-grade teachers (1999)
  - Whole school (1997)
  - Whole school (1999)

- Rating of professional development
  - All fourth-grade teachers (1998)
  - All fifth-grade teachers (1999)
  - Whole school (1998)

- Outreach to low achievers' parents
  - Whole school (1998)
  - Whole school (1998), interacting with student's initial achievement status

---

these techniques, the student was likely to gain 3.6 points more; if the teacher reported little use of them, the expected gain was 3.8 points less than average. However, for those individual students whose initial test scores were in the bottom quarter of the national distribution, the impact of this instructional variable was less strong. For such students, if their fifth-grade teacher was at the 90th percentile in the use of exploration in instruction, they were likely to gain 2.6 points more than average; if the teacher was only at the 10th percentile on this variable, they were likely to gain 2.7 points less than average.

Table 3-8. Variables used to predict each longitudinal student's third-grade score in the final HLM mathematics model

---

Control variables

- Student poverty
- School poverty
- School size in 1997

Student-level variables

- Exploration in instruction
  - Third-grade teacher (1997)
  - Third-grade teacher (1997), interacting with school poverty

School-level variables

- Visibility of standards and assessments
  - Whole school (1997)

- Exploration in instruction
  - All third-grade teachers (1997)
  - All third-grade teachers (1997), interacting with school poverty

---

A fifth-grade teacher's positive rating of professional development in mathematics was associated with 3 more points gained over 2 years. A negative rating for professional development (which, in the case of this teacher-level variable, would mean a total lack of professional development in mathematics over the past year) was associated with gains that were 3 points less than average.

In the sample's poorest schools, a fourth-grade teacher's responses to the index of visibility of standards and assessments were negatively related to his or her student's gains (with gains 2.6 points less for 90th-percentile teachers and gains 2.4 points greater for 10th-percentile teachers). In the less poor schools, the opposite relationship was found. There, a teacher who was paying more attention to standards and assessments would have students who made gains 3.1 points greater, while a teacher paying less attention to these policy instruments would have students making gains that were 2.9 points less.

The teacher's reported use of the NCTM standards also showed effects only in conjunction with school poverty, but in this case the effects were the opposite of those associated with the standards and assessments developed by states or districts. In high-poverty schools, a third-grade teacher's reported high use of NCTM standards was associated with greater student gains—4.5 points more. In less poor schools in this sample, a teacher's reported use of NCTM standards was associated with gains that were 5.4 points less than average.

Table 3-9. Final mathematics HLM model: Effects on score gains of longitudinal students, significant independent variables

Average Gains				
Points gained, third to fifth grade, under average conditions on all independent variables	46.2[a]			
Difference in 2-year gain associated with high/low value on each significant independent variable				
Control variables	**School at the 90th percentile**		**School at the 10th percentile**	
School poverty	+5.8[c]		-6.6[c]	
School's low base-year test scores	-10.8[b]		+9.4[b]	
Student-level variables	**Teacher at the 90th percentile**		**Teacher at the 10th percentile**	
Visibility of standards and assessments (fourth-grade teacher, 1998), interacting with school poverty	**And 90th percentile school poverty**	**And 10th percentile school poverty**	**And 90th percentile school poverty**	**And 10th percentile school poverty**
	-2.6[b]	+3.1[b]	+2.4[b]	-2.9[b]
Reported use of NCTM standards (third-grade teacher, 1997), interacting with school poverty	+4.5[a]	-5.4[a]	-4.8[a]	+5.7[a]
Exploration in instruction (fifth-grade teacher, 1999)	+3.6[a]		-3.8[a]	
Exploration in instruction (fifth-grade teacher, 1999), for student with low initial achievement	-1.0[b]		+1.1[b]	
Rating of professional development (fifth-grade teacher, 1999)	+3.0[a]		-3.0[a]	
School-level variables	**School at the 90th percentile**		**School at the 10th percentile**	
Visibility of standards and assessments (all fourth-grade teachers, 1998)	-7.6[a]		+7.0[a]	
Presentation and practice in instruction (all fifth-grade teachers, 1999), interacting with school poverty	**And 90th percentile school poverty**	**And 10th percentile school poverty**	**And 90th percentile school poverty**	**And 10th percentile school poverty**
	-10.2[a]	+12.2[a]	+7.8[a]	-9.4[a]
Reported use of NCTM standards (all third-grade teachers, 1997), interacting with school poverty	-14.0[a]	+16.7[a]	+10.1[a]	-12.1[a]
Preparation for instruction (whole school, 1999)	-5.2[c]		+5.5[c]	
Rating of professional development (all fourth-grade teachers, 1998)	+3.9[c]		-8.3[c]	
Outreach to low achievers' parents (whole school, 1998), for student with low initial achievement	+9.9[a]		-7.1[a]	

Table reads: For average conditions on all independent variables, the model predicts a gain of 46.2 points. Students whose school was at the 90th percentile for school poverty would gain 5.8 points more; students whose school was at the 10th percentile for poverty would gain 6.6 points less.

[a] Significant effect at the .001 level

[b] Significant effect at the .01 level

[c] Significant effect at the .05 level

- **Many instructional variables measured at the school level had significant effects on gains, either alone or in interaction with other variables.** The relative visibility of standards and assessments for all fourth-grade teachers had a negative relationship with student gains, echoing the relationship found for the student's individual fourth-grade teacher in high-poverty schools. Across the board, schools at the 90th percentile of the LESCP sample on this variable in 1998 had student gains in mathematics 7.6 points less than average, while schools at the 10th percentile had gains 7 points greater than average.

  Two instructional variables were negatively associated with the rate of gain in high-poverty schools but positively associated in the less poor schools. In high-poverty schools, relatively high levels of presentation and practice in fifth-grade teachers' instruction meant that students were gaining 10.2 points less than students whose teachers were at the average on this index. It is possible that these instructional practices were a response to the students' low levels of performance. In less poor schools, on the other hand, high levels of presentation and practice were associated with gains 12.2 points higher than average.

  Similarly, the extent to which third-grade teachers said they were using NCTM standards was associated with slower rates of score gains in high-poverty schools: in the poorest schools, the rate was 14 points less than average where the teachers said they were using these standards. In the less poor schools, however, a cadre of third-grade teachers using NCTM standards in their curriculum was associated with gains that were 16.7 points above average.

  The finding about reported use of NCTM standards by the whole third grade is, perplexingly, the opposite of the finding for individual third-grade teachers. Our best judgement about this set of findings is that teachers' reported use of NCTM standards, which was uncorrelated with any other responses on the surveys, was simply not a variable that was robust enough to show meaningful results in the study.

  Across the entire sample, the reports of all teachers in the school about their own level of skill in various aspects of mathematics teaching were negatively related to student gains on this test. Where teachers reported being well prepared in several specific instructional skills, the model predicted students' gains 5.2 points less than average. Where teachers expressed less preparation, the gains were 5.5 points more than average. It is possible that some ineffective schools had teachers who were overconfident about their preparation and, by the same token, that other teachers, who were making a transition to more effective techniques, reported a lack of confidence in their skills.

  A finding that is intuitively easier to understand is the positive relationship between the whole school's rating of professional development in mathematics and student gains. Schools where all teachers praised the quality and purposefulness of their professional development accounted for 3.9 more points gained over 2 years; schools where teachers gave negative ratings to their professional development tended to have 8.3 fewer points gained. In combination with the positive relationship found for individual teachers, growth in test scores between grades three and five was 50 percent higher for those students whose teachers and schools rated professional development high than when they gave it a low rating.

Finally, the variable of outreach to the parents of low achievers was positively related to student gains for those students who started out in the bottom quarter of the national distribution for this test. The whole school's level of outreach to parents predicted an additional gain of 9.9 points at the 90th percentile, compared with 7.1 fewer points gained at the 10th percentile. This represents a 40-percent higher rate of score growth for students in schools whose teachers reported high levels of parental outreach than those students in schools whose teachers reported low levels of parental outreach activities.

We next report on the associations between the variables listed in Table 3-8 and longitudinal students' third-grade test scores in mathematics. Table 3-10 shows the differences in third-grade score for students whose teachers or schools met particular conditions measured by our indices, holding all other variables at their average in the model. It shows only the effects of those variables that had significant effects in the final model, dropping a number of variables that were shown in Table 3-8 but that proved not to be significantly related to student achievement in this final multivariate analysis.

Findings for Third-Grade Mathematics Scores

- **Similar to the associations found in reading, poverty showed a significant and negative association with third-grade mathematics achievement.** Students eligible for free or reduced-price lunch would score 5.8 points lower than the predicted average score, which was 597.1; students not eligible for the lunch program would score 7.5 points higher. Attending a school with very high or low poverty also had a statistically significant association with initial achievement: in a school at the 90th percentile of this sample for poverty, a student would score 8.9 points lower than the average, all other things at the average, while a student in a comparatively less impoverished school (for this sample) would score 10.2 points higher.

- **Only one student-level variable, exploration in instruction, was related to initial achievement, and only in interaction with school poverty.** In high-poverty schools, a third-grade teacher's greater frequency and duration of these student-centered, exploratory activities was associated with higher student performance. In such a school, a teacher at the 90th percentile on this instructional variable would have students with third-grade scores 3.9 points above the average, while a teacher at the 10th percentile would have students with scores 3 points below the average. The opposite relationship prevailed in less impoverished schools, where the teacher's relatively high use of exploration in instruction was associated with scores 4.7 points below average, and teacher's relatively low use of this approach was associated with scores 3.6 points above average. Compounding the interaction effects is the original 8.9-point deficit associated with being in a high-poverty school.

- **Greater visibility of standards and assessments for all teachers, K–5, was associated with higher initial student scores in mathematics.** Schools at a high level on this index had students scoring 6.5 points above average, in contrast to scores 6 points below average in those schools reporting relatively low visibility of standards and assessments in mathematics.

Again, it may be helpful to review the distribution of teacher scores on these variables so as to give more concrete meaning to the relatively high or low values of particular indices that were significantly related to student achievement.

Table 3-10. Final mathematics HLM model: Effects on third-grade achievement of longitudinal students, significant independent variables only

Average third-grade score				
Third-grade score of a student experiencing average conditions on all independent variables	597.1[a]			
Difference in score associated with high/low value on each significant independent variable				
Control variables	**Student eligible for free or reduced-price lunch or school at the 90th percentile**		**Student not eligible for free or reduced-price lunch or school at the 10th percentile**	
Student poverty	-5.8[a]		+7.5[a]	
School poverty	-8.9[b]		+10.2[b]	
Student-level variables	**Teacher at the 90th percentile**		**Teacher at the 10th percentile**	
Exploration in instruction (third-grade teacher) interacting with school poverty	**And 90th percentile school poverty**	**And 10th percentile school poverty**	**And 90th percentile school poverty**	**And 10th percentile school poverty**
	+3.9[b]	-4.7[b]	-3.0[b]	+3.6[b]
School-level variables	**School at the 90th percentile**		**School at the 10th percentile**	
Visibility of standards and assessments (whole school)	+6.5[b]		-6.0[b]	

Table reads: For longitudinal students with average values on all independent variables, the model predicts a third-grade score of 597.1 on the SAT-9 closed-ended mathematics test. Students eligible for free or reduced-price lunch would score 5.8 points less; students not eligible for free or reduced-price lunch would score 7.5 points more.

[a] Significant effect at the .001 level

[b] Significant effect at the .01 level

The mean score on our Visibility of Standards and Assessments Index in 1999 was a relatively high 27.7 on a scale of 8 to 32, implying that the average teacher in our sample reported a strong familiarity with mathematics standards and assessments, as well as a high level of integration of these into the curriculum. Scores on this index were high, with little variation. Teachers who fell at the 10th percentile of the index had a score of 23, well above the possible minimum of 8, which suggested standards and assessments were moderately familiar and influential in their teaching. Teachers at the 90th percentile (with the maximum index score of 32), whom we rated as having high visibility of standards and assessments, were very familiar with the mathematics assessments and reported teaching a curriculum that was very reflective of the standards.

The mean score of 11.5 (on a scale of 4 to 16) on our Outreach to Low Achievers' Parents Index suggests that the average teacher in our sample reported initiating contact with many parents of their low-achieving students through face-to-face meetings, telephone calls, or by sending home materials. Teachers who fell at the 90th percentile had the maximum index score of 16, meaning that they reported initiating communication with the parents of most of their low-achieving students. In contrast, the score of 8 for teachers at the 10th percentile indicates contact with fewer, but still some, parents of low-achieving students.

The average teacher in our sample used activities such as manipulatives, group discussions, or long-term projects for a relatively small number of mathematics lessons, as suggested by the mean score of 286.9 out of the maximum 1,215 lessons per year for our Exploration in Instruction Index. Teachers who we reported as having relatively high levels of exploration in instruction, at the 90th percentile of the index with a score of 528.8, reported a moderate amount of these activities in their classrooms, while teachers at the 10th percentile had a score of 97.1 (compared with the minimum possible of 0).

Teachers' Rating of Professional Development in Mathematics Index displayed a large amount of variation. The average score for the index was 21.6 on a scale from 0 to 41, with a 90th percentile score of 38 and a 10th percentile score of 0. This indicates that teachers at the 90th percentile, who we considered to have a high rating of professional development, reported a very high quality of mathematics professional development activities, as well as professional development that matched standards, assessments, and reform plans and that enhanced their skills and knowledge. In contrast, teachers who fell at the low end of the index had participated in no professional development in mathematics over the past year.

The average teacher in our sample reported a low frequency and/or intensity of mathematics teaching activities such as lectures, tests, worksheets, or drills, reflected in the fairly low mean of 362.7 (on a scale of 0 to 945 lessons per year) on our Presentation and Practice in Instruction Index. Teachers who we reported as having a high level of presentation and practice in instruction were at the 90th percentile of the index (score 576). This score suggests that they reported fairly low amounts of these instructional activities, although the amounts were high when compared with the practices of teachers at 10th percentile (index score 182.3).

We again looked for students in the sample who came closest to experiencing the combination of favorable instructional conditions. Forty-four students experienced values above the LESCP average on the instructional variables positively related to student gains in mathematics and values below the LESCP average on the ones negatively related to gains. Their individual poverty was like that of the sample as a whole, but they were more apt to be in schools with poverty rates above 90 percent (59% of them, compared with 31% of the rest of the sample). Their performance on the closed-ended, mathematics test was as shown in Table 3-11.

Table 3-11. Mathematics base-year scores and gains for longitudinal students with favorable instructional conditions, average longitudinal students, and national norms students

	Longitudinal students with favorable instructional conditions	Average predicted by model for longitudinal sample	National norms, 1995 (50th percentile students)
Third-grade mathematics score	579	597.1	599
Two-year gain	57	46.2	47
Fifth-grade mathematics score	636	643.3	646

They began the study 18 points behind their counterparts in the LESCP sample and 20 points below the norm of 599. Their 2-year gain exceeded by 10.8 points that predicted for the longitudinal sample, which was 46.2 points. Their gain exceeded by 10 points the difference between the third- and fifth-grade national norms. Thus, these students closed 40 percent of the gap between themselves and the longitudinal sample and closed about one-half of the gap between their scores and national norms.

Summary

The HLM method enabled us to find instructional conditions that were associated with higher third-grade scores or greater gains from third grade to fifth grade, in reading or mathematics, for the students whom this study followed over 3 school years. The depth of information available about these students' classrooms and schools permits us to say, in detail, what seemed to help them end up at higher levels of performance by the end of fifth grade. The model explained a good deal of the variation in outcomes between schools in both subjects, especially with respect to students' gains.

A thumbnail sketch of the HLM findings, below, indicates what was a positive direction for each significant control and instructional variable. The summary begins with all longitudinal students (i.e., presents the main effects of the model) in reading, then attends separately to the students in special circumstances, such as very high-poverty schools. It follows a similar structure for mathematics.

Variables with **positive** relationships to **reading score gains** for **all** longitudinal students were the following:

- Less use of basic instruction by fifth-grade teacher;
- Third-grade teacher who contacted low achievers' parents;
- Higher rating of professional development in reading, schoolwide; and
- A group of third-grade teachers who contacted low achievers' parents.

There were no interaction effects on reading score gains.

Variables with **positive** relationships to **initial reading score** for **all** longitudinal students were the following:

- Family income too high to qualify for free or reduced-price lunch;
- Relatively few poor longitudinal students in the school; and
- Higher visibility of standards and assessments for the third-grade teacher.

Additionally, we found a negative relationship among teachers contacting the parents of low-achieving students and initial achievement. We interpret this to mean that third-grade teachers are doing a good job in identifying the lowest achievers in reading and then working with these students' parents. A variable with an **additional, positive** relationship to **initial reading scores** in the **poorest schools** was high visibility of standards and assessments, schoolwide.

Variables with **positive** relationships to **mathematics score gains** for **all** longitudinal students were the following:

- A high proportion of poor longitudinal students in the school;
- Fewer students in the school with low mathematics scores in third grade;
- Considerable time spent in exploration in fifth-grade mathematics instruction;
- Higher rating of professional development in mathematics by the fifth-grade teacher;
- Relatively low visibility of standards and assessments between all fourth-grade teachers;
- Less teacher confidence about his or her own preparation for teaching mathematics, schoolwide; and
- Higher ratings of professional development in mathematics, schoolwide.

Variables with **additional, positive** relationships to **mathematics score gains** in the **poorest schools** were the following:

- Lower visibility of standards and assessments in mathematics for the fourth-grade teacher; and
- Less time spent in presentation and practice in mathematics instruction in all fifth-grade classes.

Finally, variables with **additional, positive** relationships to **mathematics score gains** for students with initially **low scores** were the following:

- Less time spent in exploration in fifth-grade mathematics instruction; and
- Higher rates of outreach to low achievers' parents, schoolwide.

Variables with **positive** relationships to **initial mathematics score** for **all** longitudinal students were the following:

- Family income too high to qualify for free or reduced-price lunch;
- Relatively few poor longitudinal students in the school; and
- High visibility of standards and assessments, schoolwide.

The variable with an **additional, positive** relationship to **initial mathematics scores** in the **poorest schools** was considerable time spent in exploration in third-grade mathematics instruction.

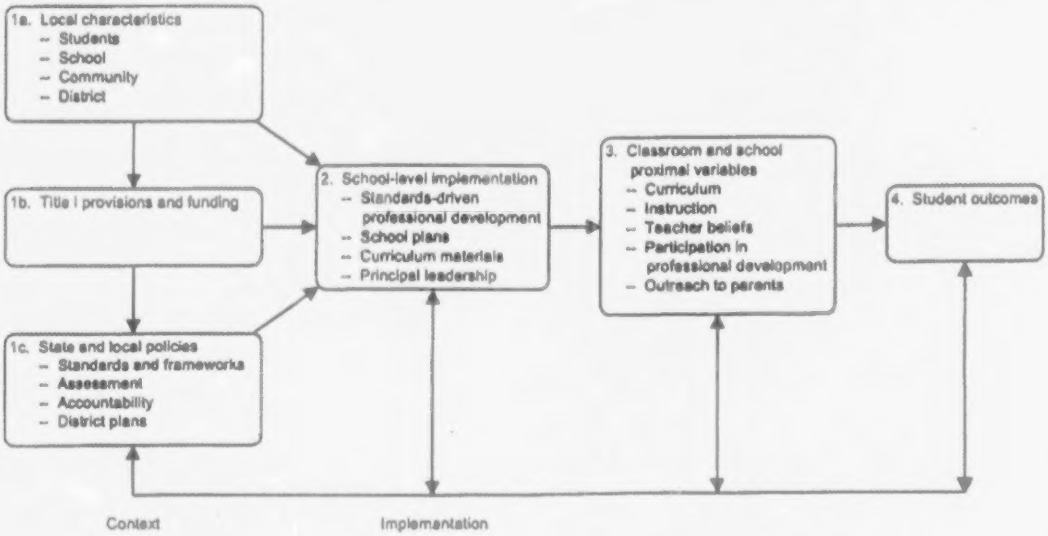
An examination of students who experienced above average values on the LESCP variables that had a significant, positive relationship to score gains revealed the following:

- In reading, these students were disproportionately poor, and so were their schools;
- In reading, having started out in the third grade 10.6 points behind their peers and 18 points below national norms, they closed the gap to end up 4.9 points behind their peers and 12 points below national norms by the end of the fifth grade;
- In mathematics, these students' rate of poverty was similar to others in the LESCP study, but they attended schools with higher poverty rates; and
- In mathematics, these students were 18.1 points behind their peers and 20 points below the national norm at the end of the third grade, but only 7.3 points behind their peers and 10 points below the national norm at the end of the fifth grade.

CONTEXT AND IMPLEMENTATION VARIABLES RELATED TO CLASSROOM AND SCHOOL-LEVEL PRACTICES

Having identified conditions in classrooms and schools that seemed to make a difference for student outcomes over the course of the study, we turn now to the larger arena of school reform. To do this, we introduce the boxes at the left-hand side of the conceptual model, repeated here as Figure 4-1. First, we look at the relationship between some of the local characteristics on the one hand (Box 1a of Figure 4-1) and the key instructional variables on the other (Box 3 of Figure 4-1), addressing the issue of whether poor students, failing students, or high-poverty schools were getting access to good instruction or strong instructional support. We pay particular attention here to the students and schools that are the traditional concern of Title I policy: students with low achievement in Title I schools and schools with high levels of poverty. Second, we ask what features of the policy environment (Boxes 1b, 1c, and 2 of Figure 4-1) may have stimulated and supported the instructional conditions that mattered for students. Thus, we make the link back to standards-based reform, asking whether the states and school districts that were doing the most to encourage reform, compared with others in the Longitudinal Evaluation of School Change and Performance (LESCP) sample, had succeeded in creating more favorable instructional conditions for their students.

Figure 4-1. Conceptual framework


1a. Local characteristics
– Students
– School
– Community
– District
1b. Title I provisions and funding
1c. State and local policies
– Standards and frameworks
– Assessment
– Accountability
– District plans
2. School-level implementation
– Standards-driven professional development
– School plans
– Curriculum materials
– Principal leadership
3. Classroom and school proximal variables
– Curriculum
– Instruction
– Teacher beliefs
– Participation in professional development
– Outreach to parents
4. Student outcomes
Context
Implementation

BEST COPY AVAILABLE

Specifically, we address four key issues in this chapter:

- Are the poorest students getting access to favorable instructional conditions?
- Are the students who are initially the lowest achievers getting access to favorable instructional conditions?
- Are students in the highest poverty schools getting access to favorable instructional conditions?
- Are favorable instructional conditions related to state and school district implementation of standards-based reform?

Favorable instructional conditions are defined here to be the main effects determined to have a statistically significant relationship with either third-grade test score or third- to fifth-grade test score growth, from the hierarchical linear modeling (HLM) analyses described in Chapter 3. A summary of these conditions is shown in Table 4-1. A positive sign in a cell indicates that more of the practice has a positive effect on third-grade test score or third- to fifth-grade growth. A negative sign in a cell indicates that less of the practice was favorable. Blank cells indicate no statistically significant main effect relationship with student achievement.

Poor and Initially Low-achieving Students' Access to Favorable Instructional Conditions

Reading

Table 4-2 summarizes the results of the analyses of differential access to favorable reading instructional conditions by the following:

- Student poverty (as measured by whether or not the student was eligible for free or reduced price lunch),
- Student initial achievement status (bottom quarter nationally versus top three-quarters nationally on the third-grade Standard Achievement Test, Ninth Edition (SAT-9)), and
- School poverty concentration (as measured by the percentage of students in the school who were eligible for free or reduced-price lunch).

Table 4-1. Direction of the main effects of teaching practices from the HLM analyses

Reading	Third-grade practice on third- to fifth-grade gain	Fourth-grade practice on third-to fifth-grade gain	Fifth-grade practice on third- to fifth-grade gain
Visibility of standards and assessments			
Basic instruction in the upper grades			-
Preparation for instruction			
Rating of professional development			+
Outreach to low achievers' parents	+		

Mathematics	Third-grade practice on third- to fifth-grade gain	Fourth-grade practice on third-to fifth-grade gain	Fifth-grade practice on third- to fifth-grade gain
Visibility of standards and assessments		-	
Reported use of National Council of Teachers of Mathematics standards			
Exploration in instruction			+
Presentation and practice in instruction			
Preparation for instruction			-
Rating of professional development		+	+
Outreach to low achievers' parents			

Table 4-2. Differences in favorable instructional indices for reading scores by student poverty, student achievement status in 1997, and school poverty concentration

	Conditions of index associated with greater achievement	Differences in favorable conditions by		
		Student poverty	Student achievement status in 1997	School poverty concentration
Favorable effects on third- to fifth-grade reading gain				
Basic instruction in upper grades	**Less** of this practice in **fifth** grade	Students not eligible for free or reduced-price lunch received significantly less	Students who scored in the top ¾ in 1997 received significantly less	No significant difference by school poverty concentration
Rating of professional development	**More** of this practice in **fifth** grade	Students eligible for free or reduced-price lunch received significantly more	Students who scored in the bottom ¼ in 1997 received significantly more	No significant difference by school poverty concentration
Outreach to low achievers' parents	**More** of this practice in **third** grade	No significant difference by student poverty	No significant difference by student achievement status	No significant difference by school poverty concentration

BEST COPY AVAILABLE

The first column of Table 4-2 lists the conditions found by the HLM analyses to be favorable for third-grade to fifth-grade growth in reading achievement. These conditions are the main effect variables summarized in Table 4-1 for reading. The second column of Table 4-2 indicates the direction of the effect and the grade in which the practice had an effect.

The third column of Table 4-2 presents the result of the analyses of differences in these practices among students eligible for free or reduced-price lunch and those who were not eligible for free or reduced-price lunch. We analyzed the extent to which the LESCP longitudinal students from poor families, as indicated by their eligibility for free or reduced-price lunch, had the opportunity to learn from teachers whose practices and beliefs were the kind associated with better student outcomes. For each poor student, we determined the value of the favorable practice for his or her teacher. For example, because rating of professional development by the fifth-grade teacher was deemed to have a significant effect on third-grade to fifth-grade reading score growth, we calculated a value of that measure for each poor student's fifth-grade teacher. We then averaged these values over all of the poor (free or reduced-price lunch-eligible) students in the LESCP longitudinal sample. We then did the same thing for students who were not eligible for free or reduced-price lunch and compared the means for the two groups. The second row of the student poverty column, indicates that there was a statistically significant difference (at the .05 level) between the mean teacher rating of professional development and assessment between poor and nonpoor students in favor of the poorer students.

The HLM analyses demonstrated that less basic instruction in reading in the upper grades was associated with greater gains in reading scores between the third and fifth grades. The use of basic instruction in the upper grades, which was a consistent feature of poor students' experience, appears to have worked to the poorer students' detriment.

Continuing down the third column of Table 4-2, we see no difference in access to outreach to low-achieving student's parents by third-grade teachers. Thus, on average, poor and nonpoor students in the LESCP sample were taught by teachers who reported the same level of outreach to low achievers' parents. Saying this another way, on average, poor students had access to the same level of this favorable behavior as did less poor students.

Because Title I policy pays particular attention to those students in Title I schools who start out with low achievement, we looked at the instructional strategies being offered to those students, again using our HLM results as the guide to what strategies were important. Student achievement status (the fourth column of Table 4-2) was defined into two categories: the student having received a third-grade SAT-9 reading score at or below the 25th percentile nationally or having a score above the 25th percentile nationally.

The results by initial achievement status are similar to those found in the analysis by student poverty. Our findings show no overwhelming pattern of advantages or disadvantages accruing to the students with low initial scores in reading. Their teachers were more likely to focus on basic instruction. By fifth grade, this may have placed the initially low-achieving students at a disadvantage.[1] On the other hand, these students could have benefited from their teachers' higher ratings of professional development in reading.

We now shift our focus to whole schools and examine the extent to which the instructional indices may have varied with the school's poverty level. We looked at the mean for each index among all teachers, grades K to 5, in each year, by poverty level of their school.

In reading, as shown in the last column of Table 4-2, there were no significant differences favoring the schools serving fewer poor students, relative to others in the LESCP sample. In those schools, teachers reported essentially the same use of basic instruction in the upper grades and the same rating of professional development, compared with the higher poverty schools, when the LESCP longitudinal students were in the fifth grade. Likewise, the teachers reported the same average value for outreach to low achievers' parents when the students were in the third grade.

Overall, in reading, the poorer students and the lower initial achieving students tended to be exposed to the negative practice of added emphasis in basic instruction in reading in the fifth grade. However, these same students benefited from having teachers who rated high in their report of higher quality professional development. No differences in significant practices were observed between high and less high poverty schools in the LESCP sample.

---

[1] One could argue that because these students were low achievers in the third grade, basic instruction in reading might be appropriate, even in the fifth grade. However, the HLM analyses showed no significant interaction effect between basic instruction in the upper grades and initial achievement status. Thus, the HLM analysis did not indicate that these initial low achievers benefited from basic reading instruction and, therefore, basic instruction in reading in the fifth grade had an equally negative effect on the growth in reading scores for low initial achievers and high initial achievers.

Mathematics

Similar findings for mathematics appear in Table 4-3, which shows more differences but again reveals a mixture of advantages and disadvantages for poor students. For example, poor students were at an advantage from the low visibility of standards and assessments, when the lower values among their fourth-grade teachers were associated with greater rates of score gain. In fifth grade, poor students had the disadvantage of teachers who may have been overconfident about their preparation for instruction, but they had the advantage of teachers who gave high ratings to their professional development in mathematics.

Unlike reading, in mathematics, the teachers of initially low-scoring students showed no differences in comparison with the teachers of initially higher scoring students among the indices associated with favorable teaching practices. We now turn to the last column of Table 4-3, school poverty concentration.

In mathematics, the only significant difference that emerged was in the index of exploration in instruction. Teachers in the higher poverty schools reported spending more time in these more student-directed, open-ended activities. It was an advantage gained by students in higher poverty schools because this was a type of activity that was associated with greater score increases for students.

Policy Environment and Favorable Instructional Conditions

An important issue of this study is the investigation of the effects of standards-based reform on instructional conditions and, thereby, on student outcomes. We therefore analyzed the extent to which favorable instructional conditions (as identified through the HLM analysis of student performance) were occurring in those environments in the LESCP sample that had high levels of activity in particular aspects of standards-based reform.

To do this, we first categorized the states and districts participating in the study by their activities in standards-based reform. We created simple measures (yes-no or yes-partial-no) of several specific aspects of reform policy. Some centered on the measurement and reporting of student

Table 4-3. Differences in favorable instructional indices for mathematics scores by student poverty, student achievement status in 1997, and school poverty concentration

		Differences in favorable conditions by		
	Conditions of index associated with greater achievement	Student poverty	Student achievement status in 1997	School poverty concentration
Favorable effects on third- to fifth-grade mathematics gain				
Visibility of standards and assessments	**Less** of this practice in **fourth** grade	Students eligible for free or reduced-price lunch received significantly less	No significant difference by student achievement status	No significant difference by school poverty concentration
Exploration in instruction	**More** of this practice in **fifth** grade	No significant difference by student poverty	No significant difference by student achievement status	Schools with the highest poverty concentrations (90%-100%) had teachers reporting significantly more of this practice than schools with the lowest poverty concentration (<50%)
Preparation for instruction	**Less** of this practice in **fifth** grade	Students not eligible for free or reduced-price lunch received significantly less	No significant difference by student achievement status	No significant difference by school poverty concentration
Rating of professional development	**More** of this practice in **fourth** grade, and	No significant difference by student poverty	No significant difference by student achievement status	No significant difference by school poverty concentration
	More of this practice in **fifth** grade	Students eligible for free or reduced-price lunch received significantly more	No significant difference by student achievement status	No significant difference by school poverty concentration

performance, others on school accountability, still others on the enactment of standards or the criterion of standards in district or school plans. Each pertained to elementary schools; the documentary evidence was generally clear on each; and together they worked to differentiate among policy environments.

We must note at the outset that every district moved along on these reforms over the course of the study. Thus, in the case of several of the hallmarks of standards-based reform, the only way to distinguish policy environments from each other in this study was to identify those districts or states that had enacted a particular provision earlier than others, by 1997. If any of them had later rescinded any of these reform provisions, we would have changed the way we classified the state or district, but in fact none did.

Some of the evidence used in this analysis came from published sources. *Quality Counts* told us whether states had enacted content standards by 1997 and whether they had made their schools subject to rewards or sanctions for performance by 1997. Documents collected onsite in each district formed the primary basis for this analysis, however. For example, when no statewide system of rewards or sanctions existed, we looked in those documents for evidence of a local system. The local documents also provided evidence of the way in which achievement data were summarized and reported to the public. Finally, they formed the basis for an analysis of district and school plans. In all cases, no fewer than two analysts reviewed the local documents; all analysts were trained in the study procedures, and a lengthy process of multiple reviews and group discussion was used to ensure consistency in the ratings.

Two indices, partially overlapping in their contents, were created and used to distinguish the study districts[2] from one another with respect to their policy environments. As with teacher-level indices, statistical correlations were used in creating these indices. We identified those that fell at the high or low end with respect to each index. Some districts with values somewhere in the middle were excluded from this analysis, in our focus on clear extremes within the LESCP sample. The policy indices used were the following:

- **Focus on Assessment and Accountability.** This index was the sum of district-level scores on four measures having to do with the reporting and the consequences of assessments: whether schools in the state or district were subject to sanctions on the basis of student performance as of 1997, whether schools in the state or district were subject to rewards on the basis of student performance as of 1997, whether the district reported assessment data in terms of proficiency levels, and whether the district reported school and district achievement status to the public in readable and understandable formats using data disaggregated to two or more of the Title I categories by 1998. A total of 34 schools were in high districts, 20 in low districts.

---

[2] Most of the policy variables described here originated at the state level. In several cases, though, a district had made its own policy (e.g., for a system of accountability) in the absence of a state policy. In no case did we have evidence that schools were subject to overlapping frameworks from both levels—for example, one system of rewards or sanctions emanating from the state and another from the district.

- **Presence of Standards, Consequences, and Plans.** This index combined data on the enactment of policies on standards and consequences with the specific citation of these policies in both school and district improvement plans. It summarized five district-level measures: whether the state or district had enacted content standards in reading and mathematics by 1997, whether schools in the state or district were subject to sanctions on the basis of student performance as of 1997, whether schools in the state or district were subject to rewards on the basis of student performance as of 1997, whether district and school improvement plans consistently referred specifically to content standards in all 3 years, and whether district and school improvement plans consistently set specific goals linked to required assessments in all 3 years. There were 26 schools in high districts, eight in low districts.

Tables 4-4 and 4-5 display the similarities and differences across policy environments in the instructional variables found to be significantly related to student achievement in reading and in mathematics, respectively.

There was some advantage, for reading instruction, associated with being in a district that was undergoing standards-based reform in a relatively intensive way. There were no teaching practices that put the students at a disadvantage in these districts and benefits accrued to students in these districts from the increased quality of professional development, as rated by the teachers.

We next review the corresponding relationships in mathematics. As shown in Table 4-5, there were several significant relationships, and they tended to show some advantages for an environment of intense standards-based reform. A policy focus on assessment and accountability was associated with higher ratings for professional development in 1998, the year in which the LESCP longitudinal students were in the fourth grade—an instructional variable that was positively associated with growth in student achievement in mathematics.

The presence of standards, consequences, and plans was also associated with high teacher ratings of professional development. This higher presence was also associated with higher teacher ratings of visibility of standards and assessments, which, unfortunately, was associated with lower growth in mathematics scores between the third and fifth grades.

Table 4-4. Differences in favorable instructional indices for reading scores by policy environment

		Differences in favorable conditions by	
	Conditions of index associated with greater achievement	Policy focus on assessment and accountability	Presence of standards, consequences, and plans
Favorable effects on third- to fifth-grade reading gain			
Basic instruction in upper grades	**Less** of this practice in fifth grade	No significant difference among districts with high and low focus on this policy	No significant difference among districts with high and low presence of these items
Rating of professional development	**More** of this practice in fifth grade	No significant difference among districts with high and low focus on this policy	Districts with a high presence of these items had teachers reporting significantly more of this practice
Outreach to low achievers' parents	**More** of this practice in third grade	No significant difference among districts with high and low focus on this policy	No significant difference among districts with high and low presence of these items

BEST COPY AVAILABLE

Table 4-5. Differences in favorable instructional indices for mathematics scores by policy environment

		Differences in favorable conditions by	
	Conditions of index associated with greater achievement	Policy focus on assessment and accountability	Presence of standards, consequences, and plans
Favorable effects on third- to fifth-grade mathematics gain			
Visibility of standards and assessments	**Less** of this practice in fourth grade	No significant difference among districts with high and low focus on this policy	Districts with a low presence of these items had teachers reporting significantly less of this practice
Exploration in instruction	**More** of this practice in fifth grade	No significant difference among districts with high and low focus on this policy	No significant difference among districts with high and low presence of these items
Preparation for instruction	**Less** of this practice in fifth grade	No significant difference among districts with high and low focus on this policy	No significant difference among districts with high and low presence of these items
Rating of professional development	**More** of this practice in fourth grade, and	Districts with a high focus on this policy had teachers reporting significantly more of this practice	Districts with a high presence of these items had teachers reporting significantly more of this practice
	More of this practice in fifth grade	No significant difference among districts with high and low focus on this policy	No significant difference among districts with high and low presence of these items

Summary

This chapter assessed whether the best instructional circumstances identified in this study were disproportionately available to more affluent students, higher achieving students, or schools serving more affluent students within the LESCP sample. Generally speaking, we found no clear picture of advantage in any of these groups. **Poorer students, initially lower achieving students, and students in higher poverty school in the sample had, on average, equal access to favorable teaching practices.**

We also looked at the extent to which policy environments with high levels of standards-based reform, measured in two different ways, might be associated with more or less favorable instructional conditions. In general, the results of this analysis should be encouraging to advocates of these reforms. **Schools in environments of intense standards-based reform were offering their students generally equal or more favorable instructional conditions.**

SUMMARY AND CONCLUSIONS

The Longitudinal Evaluation of School Change and Performance (LESCP) investigated effects of standards-based reform policies on students, teachers, and schools participating in the Title I program. Its purposive sample of high-poverty elementary schools included many schools that had been subject to standards-based reform policies before those policies were required by provisions of the 1994 amendments to Title I. The sample also included, for comparison, some states and school districts that adopted standards-based reform more slowly. In the 71 participating schools, the LESCP study administered standardized tests in reading and mathematics to students as they moved from third grade through fifth grade. It also gathered detailed survey data from all teachers in the schools, asking about curriculum, instruction, beliefs and attitudes about their work, participation in professional development, and outreach to parents.

The study's analyses began with the investigation of student performance and the association, if any, between student performance and specific student or school conditions. The sample of students presented an overall picture of achievement that was not unusual for high-poverty schools.

A particular asset of the LESCP study was its longitudinal database on more than 1,000 individual students, with student performance data linked to teachers' survey responses. This enabled the study team to pursue a very detailed study of students' achievement in relation to the specific instructional conditions experienced by those students over time, with statistical controls for important background variables like poverty. As the students progressed from third through fifth grade, the study tracked their upward progress in scale-score points on the Standard Achievement Test, Ninth Edition (SAT-9) and analyzed their scores and gains in relation to survey data gathered from their own teachers and all other teachers in grades K to 5 in their schools.

Having identified the conditions that were favorable to student achievement in reading and mathematics, we went on to ask where those conditions were found within the LESCP sample. Here, we were addressing questions of great policy salience. Did students' access to favorable instructional conditions vary with circumstances like family poverty or the school's poverty rate? Did the policy environments with the greatest levels of activity in standards-based reform (relative to the generally high levels found in this sample) have more or less favorable instructional conditions?

Summary of Findings

We first looked at the achievement of students in the LESCP sample as measured by their scores on the SAT-9 in the third through fifth grades. These descriptive analyses showed the following:

- LESCP sample students achieved below the published mean scores for the national norm group in reading and mathematics in each grade tested. Those students who were tested in all of the 3 years of the study (third grade in 1997, fourth grade in 1998, and fifth grade in 1999) had higher scores on average than students who were not tested all 3 years. This longitudinal sample, however, still tested below national norms.
- This initial look at overall performance also showed that poverty mattered for the absolute levels of student performance, although it did not slow the rates of gain. Specifically, those students who were eligible for free or reduced-price lunch had lower scores than other students in the sample. The score gap associated with poverty generally persisted over the course of the study.
- Independent of a student's individual poverty, students in schools with relatively higher rates of poverty had lower scores than their peers in less impoverished schools.

We next examined the relationships among classroom and school-level measures that were hypothesized to directly affect student achievement. In selecting these proximal variables for analysis, the study focused on those that have been targets of recent reform efforts, especially standards-based reform. For both reading and mathematics, detailed survey questions had asked teachers about the following:

- Their awareness and use of external standards and assessments;
- What they and their students spent the most time doing in the classroom;
- How they would rate their own skill in specific instructional tasks;
- What their recent professional development (if any) had done for them; and
- What they did to contact the parents of low-achieving students.

The specific responses were combined into indices of instructional conditions and instructional supports that varied across the sample classrooms and schools.

To find out which proximal variables were associated with student achievement and how much they mattered, the analysis relied on complex, multivariate models that incorporated data on poverty and other background variables along with the data on instructional conditions and supports. Based on this analysis, the following variables were found to have significant associations with third-grade scores or gains made between third and fifth grades, in each subject:

- In reading, for the whole sample of longitudinal students (setting aside effects found in interaction with poverty):
    - Student poverty (negative association with third-grade score);
    - School poverty (negative association with third-grade score);
    - Visibility of standards and assessments (positive association with third-grade score);
    - Basic instruction in upper grades (negative association with third-grade to fifth-grade gain);
    - Rating of professional development (positive association with third-grade to fifth-grade gain); and
    - Outreach to low achievers' parents (negative association with third-grade score, positive association with third-grade to fifth-grade gain).
- In mathematics, for the whole sample of longitudinal students (setting aside effects found in interaction with poverty or interaction with initial achievement status):
    - Student poverty (negative association with third-grade score);
    - School poverty (negative association with third-grade score, positive association with third-grade to fifth-grade gain);
    - School's proportion of initially low-achieving students (negative association with third-grade to fifth-grade gain);
    - Visibility of standards and assessments (positive association with third-grade score, negative association with third-grade to fifth-grade gain);
    - Exploration in instruction (positive association with third-grade to fifth-grade gain);
    - Preparation for instruction (negative association with third-grade to fifth-grade gain); and
    - Rating of professional development (positive association with third-grade to fifth-grade gain).

Also emerging from these models was a statistical picture of the number of test-score points that could be gained or lost by a student whose experience with particular conditions was at a high or low extreme, relative to other students in the LESCP sample. The results, simplified, appear in Table 5-1. Omitted from Table 5-1 is the more detailed information about the level at which each variable was measured (individual teacher, grade level, and/or whole school) and the year in which it was measured. Because Table 5-1 combines effects across levels and years, interested readers may want to refer back to the more detailed results presented in Chapter 3. Only a few instructional variables—four in reading, four in mathematics—made a difference irrespective of other conditions like school poverty, but the potential gain associated with those instructional variables was not trivial.

In Chapter 4, we looked at the relationships between poverty and instructional conditions, third-grade test score and instructional conditions, and state and district reform policies and instructional conditions. The underlying policy questions of that chapter are (1) "Do the poorest students have access to the instructional conditions that this study identified as having a positive effect on achievement?" and (2) "Is there a relationship between state and district reform policies and the instructional conditions that this study identified as having a positive effect on achievement?"

- The poorest students and schools in this high-poverty sample were **not** at a systematic disadvantage in their instructional conditions. There were a few differences in the instructional conditions found for poor students and the poorest schools, relative to the nonpoor students and the less poor schools in the LESCP sample. However, these differences were a mixture of favorable and unfavorable ones in both reading and mathematics.

- The policy environments around schools, however, did seem to make a consistent difference for the implementation of some significant instructional variables inside the schools. Based on a review of documents, we rated the environment with respect to two somewhat different measures of standards-based reform. One was the focus on assessment and accountability (a measure that combined information about rewards and sanctions for schools with local procedures for handling assessment data). The other measure included the early enactment of content standards, rewards and sanctions for schools, and specific references to standards in district and school plans; the shorthand name for this measure was standards, consequences, and plans.

- In reading, the differences associated with higher levels of standards-based reform were slightly more favorable to growth in student achievement. Teachers rated their professional development higher in school districts that had a higher presence of standards, consequences, and plans. A higher rating of professional development, in turn, was associated with greater gains in reading scores between the third and fifth grades.

Table 5-1. Overall results of the analysis of student performance in relation to very high levels (90th percentile) of specific instructional conditions

Reading	**Estimated difference, if any, for third-grade score**	**Estimated difference, if any, in score gain, third to fifth grade**
Student poverty	-6	None
School poverty	-12	None
Percentage of initially low-achieving students	N/A	None
School size	None	None
Visibility of standards and assessments	+12 in high-poverty schools -7 in low-poverty schools	None
Basic instruction in the upper grades	None	-2
Preparation for instruction	None	None
Rating of professional development	None	+3
Outreach to low achievers' parents	-14 in high-poverty schools -7 in low-poverty schools	+8

Mathematics	**Estimated difference, if any, for third-grade score**	**Estimated difference, if any, in score gain, third to fifth grade**
Student poverty	-6	None
School poverty	-9	+6
Percentage of initially low-achieving students	N/A	-11
School size	None	None
Visibility of standards and assessments	+7	-10 in high-poverty schools -4 in low-poverty schools
Reported use of National Council of Teachers of Mathematics standards	None	-10 in high-poverty schools +11 in low-poverty schools
Exploration in instruction	+4 in high-poverty schools -5 in low-poverty schools	+4, but -1 for initially low-achieving students
Presentation and practice in instruction	None	-10 in high-poverty schools +12 in low-poverty schools
Preparation for instruction	None	-5
Rating of professional development	None	+4
Outreach to low achievers' parents	None	+10 for initially low-achieving students

Table reads: In reading, student poverty had a negative association with a student's third-grade reading score and no association with the score gain between third and fifth grade. The size of the effect in third grade was 6 scale-score points for a student eligible for free or reduced-price lunch.

- The picture was more mixed in mathematics. In that subject, higher ratings for professional development were found in environments with more activity in standards-based reform, and our student-level findings suggest that that was a positive difference. But in mathematics, standards-based reform also brought some differences less favorable to students, especially the greater visibility of standards and assessments, which in that subject had mixed effects at best on student performance.

Conclusions

This study's findings are provocative for reformers. While confirming the harmful effects of poverty on the performance of students and schools, the findings point to some instructional practices that may boost performance in reading and mathematics. They also show an association between some of those practices and the standards-based reform approach of the current Title I program.

Because of the LESCP study's breadth of inquiry, long duration, and level of detail, this report offers findings and conclusions that might be mistaken as verdicts on a program or guides to immediate action. But no single study should dictate policy or practice. Instead, in the view of the study team, LESCP offers some advances in methods and some findings that are well worth further investigation under a wider range of circumstances. Rather than providing a detailed prescription for teachers, schools, or policymakers, LESCP suggests that standards-based reform in Title I may be on a generally promising track. Within the LESCP sample of schools, which generally experienced high levels of poverty coupled with high levels of reform activity, the study provides some support for instruction that engages students in challenging activities, for professional development that teachers rate as purposeful and helpful, and for outreach to the parents of low achievers. It suggests, however, that where teachers register especially high levels of attention to outside standards and assessments, the results for students are mixed—sometimes positive and sometimes negative. Thus, the study is a reminder that extreme versions of this reform should be tempered with caution.

APPENDIX A

HIERARCHICAL LINEAR MODELING (HLM) ANALYSIS

HLMs were developed to examine the effect of proximal and control variables on student growth on the closed-ended Stanford Achievement Test, Ninth Edition (SAT-9) reading and mathematics tests between the third and fifth grades. Student growth was examined with a three-level HLM model. Level-1 (repeated measure) represents a linear individual growth model. Level-2 (student) represents the variability of each time or growth parameter among students within a school. Level-3 represents the variability among schools.

Analysis Approach

Unconditional Model

Although separate analyses were conducted for closed-ended reading and mathematics, the approach for each analysis was the same. The first stage of the analysis was to estimate an unconditional HLM model or model without any proximal or control variables. The unconditional model decomposed the variance of the outcomes within and across the levels examined (repeated measures, students, and schools) and provided estimates of the average third-grade score and learning rate for all longitudinal students. Box A-1 presents the equations for the unconditional model. (See Tables A-6 and A-7 for variables names and descriptions.)

Model with Control Variables Only

The second stage of the analysis tested the effect of student and school characteristics on the third-grade score and learning rate. For this analysis, independent variables were added to the student- and school-level HLM equations; the repeated-measure-level HLM equation, however, remained the same as it was in the unconditional model. Box A-2 presents the equations for the model conditional on student and school characteristics.

Box A-1. Equations for the unconditional model

Level 1-Repeated Measure

The repeated-measure level represents a linear individual growth model, which can be written as

$$Y_{tij} = \Pi_{0ij} + \Pi_{1ij}(\mathrm{RYEAR})_{tij} + e_{tij}$$

where Y_{tij} is the outcome score at time t for student i in school j, Π_{0ij} is the expected third-grade score of student ij (when RYEAR = 0), Π_{1ij} is the learning rate per year for student ij from 1997 to 1999 (where RYEAR = 1 for 1998 and 2 for 1999), and e_{tij} is a student and time specific residual.

Level 2-Student

The student level represents the variability of each growth parameter among students within a school. The equations for level 2 in the unconditional model are

$$\Pi_{0ij} = \beta_{00j} + r_{0ij} \text{ and}$$

$$\Pi_{1ij} = \beta_{10j}\,,$$

where β_{00j} is the mean third-grade score within school j and β_{10j} is the mean learning rate per year within school j.

Level 3-School

The school level represents the variability among schools. The equations for level 3 in the unconditional model are

$$\beta_{00j} = \gamma_{000} + u_{00j} \text{ and}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}\,,$$

where γ_{000} is the overall mean third-grade score (across students and schools) and γ_{100} is the overall mean learning rate per year (across students and schools).

Combined model

Combining the equations from the three levels yields the following equation:

$$Y_{tij} = \gamma_{000} + \gamma_{100}(\mathrm{RYEAR})_{tij} + r_{0ij} + u_{00j} + u_{10j}(\mathrm{RYEAR})_{tij} + e_{tij}$$

BEST COPY AVAILABLE

Box A-2. Equations for the model with the control variables only

Level 1-Repeated Measure

$$Y_{tij} = \Pi_{0ij} + \Pi_{1ij}(RYEAR)_{tij} + e_{tij} ,$$

where Y_{tij} is the outcome score at time t for student i in school j, Π_{0ij} is the expected third-grade score of student ij (when RYEAR = 0), Π_{1ij} is the learning rate per year for student ij from 1997 to 1999 (where RYEAR = 1 for 1998 and 2 for 1999), and e_{tij} is a student and time specific residual.

Level 2-Student

$$\Pi_{0ij} = \beta_{00j} + \beta_{01j}(CFRLH980)_{ij} + r_{0ij} \text{ and}$$

$$\Pi_{1ij} = \beta_{10j} + \beta_{11j}(CFRLH980)_{ij} + \beta_{12j}(LO\_NAT97)_{ij} ,$$

where

β_{00j} = mean third-grade score within school j for students of average poverty;

β_{01j} = difference in third-grade score in school j for a one-unit change in student poverty;

β_{10j} = mean learning rate per year within school j for students of average poverty who scored in the top ¾ in 1997

β_{11j} = difference in learning rate in school j for a one-unit change in student poverty; and

β_{12j} = difference in learning rate in school j for a student scoring in the bottom ¼ in 1997.

Level 3-School

$$\beta_{00j} = \gamma_{000} + \gamma_{001}(MFRLH980)_j + \gamma_{002}(ENRL\_970)_j + u_{00j} ;$$

$$\beta_{01j} = \gamma_{010} ;$$

$$\beta_{10j} = \gamma_{100} + \gamma_{101}(MFRLH980)_j + \gamma_{102}(AVGENRL0)_j + \gamma_{103}(MLONT970)_j + u_{10j} ;$$

$$\beta_{11j} = \gamma_{110} \text{; and}$$

$$\beta_{12j} = \gamma_{120} ;$$

Box A-2. Equations for the model with the control variables only (continued)

where

γ_{000} = mean third-grade score for students of average poverty in schools of average size and poverty;

γ_{001} = difference in mean third-grade score for a one-unit change in school poverty;

γ_{002} = difference in mean third-grade score for a one-unit change in school size;

γ_{010} = difference in mean third-grade score for a one-unit change in student poverty;

γ_{100} = mean learning rate per year for students of average poverty who scored in the top ¾ in 1997 and attended schools of average size, poverty, and 1997 achievement;

γ_{101} = difference in learning rate for a one-unit change in school poverty;

γ_{102} = difference in learning rate for a one-unit change in school size;

γ_{103} = difference in learning rate for a one-unit change in percentage of students in a school that scored in the bottom ¼ in 1997;

γ_{110} = difference in learning rate for a one-unit change in student poverty; and

γ_{120} = difference in learning rate for a student being in the bottom ¼ nationally in 1997.

Combined model

$$\begin{aligned}Y_{tij} = {} & \gamma_{000} + \gamma_{001}(\text{MFRLH980})_j + \gamma_{002}(\text{ENRL\_970})_j + \gamma_{010}(\text{CFRLH980})_{ij} + \gamma_{100}(\text{RYEAR})_{tij} \\ & \gamma_{101}(\text{MFRLH980})_j(\text{RYEAR})_{tij} + \gamma_{102}(\text{AVGENRL0})_j(\text{RYEAR})_{tij} + \gamma_{103}(\text{MLONT970})_j(\text{RYEAR})_{tij} + \\ & \gamma_{110}(\text{CFRLH980})_{ij}(\text{RYEAR})_{tij} + \gamma_{120}(\text{LO\_NAT97})_{ij}(\text{RYEAR})_{tij} + \\ & r_{0ij} + u_{00j} + u_{10j}(\text{RYEAR})_{tij} + e_{tij}\end{aligned}$$

All student-level control variables were centered, except for a dummy flag representing whether the student scored in the bottom ¼ nationally in 1997 (LO_NAT97). Student-level variables were centered as follows:

Step 1. The grand mean of the variable was calculated across all longitudinal students;

Step 2. The student's individual value was subtracted from the grand mean to create a grand-mean-centered student value;

Step 3. The mean of the grand-mean-centered student value was calculated for each school; and

Step 4. The student-level-centered variable equaled the student's grand-mean-centered value (from step 2) minus the mean of the grand-mean-centered student value for the student's school (from step 3).

The bottom ¼ flag was not centered because it was later used as an interaction term.

All school level variables, except for school size in 1997 and average school size from 1997 to 1999, were centered to represent the mean of the grand-mean-centered student value for each school (from step 3 above). Because the school enrollment variables were based on all students in the school, not just the longitudinal students, these variables were grand mean centered across all schools.

Student- and school-level poverty were based on whether a student was eligible to receive free or reduced-price lunch in 1998. An examination of student eligibility for free lunch in 1997, 1998, and 1999 indicated that this characteristic did not change considerably from year to year. As a result, we used the information from the middle year of the analysis. If any student was missing on this item in 1998, but had information from another year, the available information was used.

Models with Control Variables and One Proximal Variable at a Time

The next stage of the analysis involved adding a single teaching practice to the model with the student and school characteristics. For each longitudinal student, we have his third-, fourth-, and fifth-grade teacher's instructional practices.[1] Furthermore, for each year of the study, we have two school-level values for each instructional practice. The first is the average value of all teachers of longitudinal students in a school in a given grade in a given year.[2] For example for 1997, we have a school-level average of an instructional practice that is based on the responses of all third-grade teachers who taught a longitudinal student. For 1998, this value is based on the responses of all fourth-grade teachers of longitudinal students. The second school-level value is the average of all teachers in a school in a given year. For example for 1997, we have a school-level average of an instructional practice that is based on all teachers in the school (K–5).[3] The student-level teaching practice and the school average based on the teachers of longitudinal students were centered using the method described in the previous section. The variables based on all teachers were grand mean centered across schools.

---

[1] Any missing data items were imputed using the hot deck imputation method.

An important question for this analysis was whether the effect of a teaching practice was conditional on school poverty or students' previous achievement. As a result, two two-way interaction terms were created for each teaching practice to investigate differential effects by (1) the extent of student poverty in a school (based on the longitudinal students) and (2) whether the student scored in the bottom ¼ nationally on the closed-ended test under investigation in the third grade.

For each teaching practice in a given grade or year, we ran four models. Model 1 included the control variables and the student-level and two school-level versions of a teaching practice index for a given year or grade. Model 2 included these variables plus the two-way interactions of the teaching practice and whether the student scored nationally in the bottom ¼ in the third grade. Model 3 replaced the bottom ¼ interactions with the two-way school poverty interactions. Model 4 included both sets of two-way interactions. No three-way interactions were included in the model. Table A-1 identifies which variables were included in each of the four models.

We were mindful of potential problems associated with multicollinearity among the various versions of a single teaching practice included in one model. As a result, we calculated correlations among the coefficients of these parameters. If the absolute value of any correlation was higher than 0.80, one of the parameters was dropped and the model was run again.

---

[2] We include this school-level variable in the model as a consequence of the centering method described in the previous section. This school-level variable comes from step 3 (i.e., the school mean of the grand-mean-centered student value).

[3] No imputations were performed for missing data on the school averages based on all teachers.

Table A-1. Variables included in models testing the effect of one teaching practice at a time, by model

		Teaching practice			Teaching practice interacted with bottom ¼ flag			Teaching practice interacted with school poverty		
Model	Control variables	Student level[a]	School level I[b]	School level II[c]	Student level[a]	School level I[b]	School level II[c]	Student level[a]	School level I[b]	School level II[c]
Model 1	X	X	X	X						
Model 2	X	X	X	X	X	X	X			
Model 3	X	X	X	X				X	X	X
Model 4	X	X	X	X	X	X	X	X	X	X

[a] Student-level represents the teaching practice of a student's individual teacher.

[b] School-level I represents the average teaching practice of teachers of longitudinal students in a school.

[c] School-level II represents the average teaching practice of all K-5 grade teachers in a school.

Note: An "X" indicates that the variable was included in the model.

BEST COPY AVAILABLE

We used the Schwarz Bayesian Criterion (BIC) to determine which of the four models had the "best fit" for each teaching practice in a given year. Teaching practices that were significant at the 0.05 level or beyond in the best fitting model were retained for the final model. In addition, because there were strong theoretical arguments for the interactions tested, any interaction that was not in the best fitting model but reached a moderate level of significance (0.01 or better) was retained for the final model.

Final Models

The primary variables included in the final reading and mathematics models were the teaching practices that were significant in the best fit model described above and the control variables. A few variables that were not significant in earlier runs were also included in the final models because of their relationship with other significant variables. For example, if a student-level teaching practice was significant, we included its corresponding school-level version that was based on the teachers of longitudinal students, even if the school-level variable was not significant. This was done because the student-level variables were group centered (i.e., within the school) and the school-level versions of the variables, based on teachers of longitudinal students, permits the model to take into account variation across schools. Additionally, if an interaction term was significant, we included its first-order terms in the final model, even if they were not significant.

HLM Results for Reading

Table A-2 presents the parameter estimates and associated standard errors from the unconditional model for closed-ended reading. Table A-3 presents these values from the model with the control variables only. Table A-4 summarizes the results from the analysis of teaching practices one at a time, 1 year at a time. It identifies which teaching practices in which years had a significant effect on the third-grade score and/or student growth according to the best fitting models. Table A-4 also identifies interaction terms that had a significant effect at the 0.01 level in models other than those with the best fit. Specifically, for those teaching practices that had a significant effect, Table A-4 presents the direction and strength of the effect, in what year or grade the effect occurred, and at what level the effect occurred (i.e., student or school). Finally, Table A-5 presents the parameter estimates and associated standard errors from the final reading model. (See Tables A-6 and A-7 for variable names and descriptions.)

Table A-2. HLM reading effects unconditional model

Fixed effects

	Gamma coefficients	Standard error	*t*-Statistic	*p*-Value
Intercept	608.20	2.33	261.27	<.0001
Slope (RYEAR)	19.48	0.58	33.78	<.0001

Random effects

Parameter	Variance estimate	Standard error	Z value	*p*-Value
Intercept across schools	286.13	63.99	4.47	<.0001
Slope across schools	12.86	3.67	3.50	0.0002
Intercept within schools	874.89	38.01	23.02	<.0001
Within student (residual)	309.10	8.35	37.02	<.0001

Goodness of fit

Criteria	Value
AIC (# of parameters)	-19688 (2)
BIC (# of parameters)	-19696 (2)

Note: Significant effects are shaded in gray.

Table A-3. HML reading effects model with control variables only

Fixed effects

	Gamma coefficients	Standard error	*t*-Statistic	*p*-Value
Intercept	607.44	1.94	313.07	<.0001
MFRLH980	-42.61	8.57	-4.97	<.0001
CFRLH980	-13.73	2.22	-6.18	<.0001
ENRL_970	0.00	0.01	-0.35	0.7305
Slope (RYEAR)	19.30	0.62	31.26	<.0001
MFRLH980*RYEAR	-1.07	2.80	-0.38	0.7031
CFRLH980*RYEAR	0.38	0.83	0.46	0.6456
RYEAR*AVGENRL0	0.00	0.00	-0.07	0.9420
RYEAR*MLONT970	1.22	3.42	0.36	0.7212
RYEAR*LO_NAT97	0.56	0.74	0.75	0.4519

Random effects

Parameter	Variance estimate	Standard error	Z value	*p*-Value
Intercept across schools	175.20	43.52	4.03	<.0001
Slope across schools	12.34	4.02	3.07	0.0011
Intercept within schools	855.23	40.07	21.34	<.0001
Within student (residual)	307.51	8.59	35.79	<.0001

Goodness of fit

Criteria	Value
AIC (# of parameters)	-19662 (10)
BIC (# of parameters)	-19678 (10)

Note: Significant effects are shaded in gray.

Table A-4. Level, year/grade, and direction of effect for teaching practices with significant effects when examined one at a time, 1 year at a time, for the best fit models

Teaching practice	Significant effect on third-grade score			Significant effect on learning rate		
	Student level[a]	School level I[b]	School level II[c]	Student level[a]	School level I[b]	School level II[c]
Basic instruction in upper grades				Fifth grade (-)		
Visibility of standards and assessments	Third grade (+)		1997 interaction with school poverty (++)[d]			1998 (+)
Rating of professional development						1999 (+)
Outreach to low achievers' parents	Third grade (—) & Third-grade interaction with school poverty (–)[d]			Third grade (+++)	Third grade (++)	
Preparation for instruction			1997 interaction with school poverty (++)[d]			

[a] Student level represents the teaching practice of a student's individual teacher.

[b] School level I represents the average teaching practice of teachers of longitudinal students in a school.

[c] School level II represents the average teaching practice of all K–5 grade teachers in a school.

[d] Interaction was not in best fitting model but was significant at the 0.01 level in another model.

Sign	Meaning
+	A positive significant effect was detected at the 0.05 level.
++	A positive significant effect was detected at the 0.01 level.
+++	A positive significant effect was detected at the 0.001 level.
-	A negative significant effect was detected at the 0.05 level.
–	A negative significant effect was detected at the 0.01 level.
—	A negative significant effect was detected at the 0.001 level.

Table A-5. HLM reading effects final model conditional on student and school characteristics and teaching practices

Fixed effects				
	Gamma coefficients	Standard error	*t*-Statistic	*p*-Value
Intercept	608.56	1.77	343.33	0.000
MFRLH980	-35.12	8.03	-4.38	0.000
CFRLH980	-13.35	2.22	-6.03	0.000
ENRL_970	-0.02	0.01	-1.49	0.142
CINF_970	0.86	0.31	2.80	0.005
MINF_970	0.91	0.81	1.12	0.267
MSF_970	1.04	1.38	0.76	0.452
MFRLH980*MSF_970	12.51	4.71	2.66	0.010
CINI_970	-1.69	0.42	-4.05	0.000
MFRLH980*CINI_970	-3.47	1.50	-2.32	0.020
MINI_970	0.01	0.99	0.01	0.989
MFRLH980*MINI_970	0.75	4.12	0.18	0.856
MSK_970	-1.10	2.45	-0.45	0.656
MFRLH980*MSK_970	-4.51	11.91	-0.38	0.706
Slope (RYEAR)	19.14	0.56	33.93	0.000
MFRLH980*RYEAR	-1.36	2.70	-0.50	0.615
CFRLH980*RYEAR	0.59	0.83	0.71	0.480
RYEAR*AVGENRL0	0.00	0.00	-1.02	0.306
RYEAR*MLONT970	5.99	3.29	1.82	0.069
RYEAR*LO_NAT97	0.63	0.75	0.84	0.402
RYEAR*CID2A990	-0.01	0.01	-2.47	0.014
RYEAR*MID2A990	-0.01	0.01	-1.80	0.072
RYEAR*MSF_980	0.45	0.33	1.37	0.171
CINI_970*RYEAR	0.76	0.16	4.87	0.000
MINI_970*RYEAR	0.82	0.26	3.17	0.002
RYEAR*MSH2A990	0.41	0.16	2.49	0.013

Random effects

Parameter	Variance estimate	Standard error	Z value	*p*-Value
Intercept across schools	124.76	34.30	3.64	<.0001
Slope across schools	7.74	2.67	2.90	0.0019
Intercept within schools	837.95	39.61	21.16	<.0001
Within student (residual)	300.53	8.47	35.49	<.0001

Goodness of fit

Criteria	Value
AIC (# of parameters)	-19226 (26)
BIC (# of parameters)	-19260 (26)

Note: Significant effects are shaded in gray.

BEST COPY AVAILABLE

Table A-6. Names and descriptions of variables used to predict each longitudinal student's third-grade score in the final HLM reading model

	Variable name
Control variables	
■ Student poverty	CFRLH980
■ School poverty	MFRLH980
■ School size in 1997	ENRL_970
Student-level variables	
■ Visibility of standards and assessments	
- Third-grade teacher (1997)	CINF_970
■ Outreach to low achievers' parents	
- Third-grade teacher (1997)	CINI_970
- Third-grade teacher (1997), interacting with school poverty	MFRLH980*CINI_970
School-level variables	
■ Visibility of standards and assessments	
- All third-grade teachers (1997)	MINF_970
- Whole school (1997)	MSF_970
- Whole school (1997), interacting with school poverty	MFRLH980*MSF_970
■ Preparation for instruction	
- Whole school (1997)	MSK_970
- Whole school (1997), interacting with school poverty	MFRLH980*MSK_970
■ Outreach to low achievers' parents	
- All third-grade teachers (1997)	MINI_970
- All third-grade teachers (1997), interacting with school poverty	MFRLH980*MINI_970

Table A-7. Names and descriptions of variables used to predict each longitudinal student's learning rate in the final HLM reading model

	Variable name
Control variables	
■ Student poverty	CFRLH980*RYEAR
■ School poverty	MFRLH980*RYEAR
■ Average school enrollment	RYEAR*AVGENRL0
■ Student's initial achievement status	LO_NAT97*RYEAR
■ School's initial achievement	MLONT970*RYEAR
Student-level variables	
■ Basic instruction in upper grades	
- Fifth-grade teacher (1999)	RYEAR*CID2A990
■ Outreach to low achievers' parents	
- Third-grade teacher (1997)	CINI_970*RYEAR
School-level Variables	
■ Visibility of standards and assessments	
- Whole school (1998)	RYEAR*MSF_980
■ Basic instruction in upper grades	
- All fifth-grade teachers (1999)	RYEAR*MID2A990
■ Rating of professional development	
- Whole school (1999)	RYEAR*MSH2A990
■ Outreach to low achievers' parents	
- All third-grade teachers (1997)	MINI_970*RYEAR

HLM Results for Mathematics

The findings from the HLM analysis of student growth on closed-ended mathematics are found in Tables A-8 to A-11. The parameter estimates and standard errors from the unconditional model are presented in Table A-8. These values for the model with just the control variables are presented in Table A-9. Table A-10 summarizes the results of the analysis of teaching practices one at a time, 1 year at a time, while Table A-11 presents parameter estimates and standard errors from the final mathematics model. Tables A-12 and A-13 provide variable names and descriptions.

Table A-8. HLM mathematics effects unconditional model

Fixed effects

	Gamma coefficients	Standard error	*t*-Statistic	*p*-Value
Intercept	598.22	2.36	253.15	<.0001
Slope (RYEAR)	22.39	0.92	24.41	<.0001

Random effects

Parameter	Variance estimate	Standard error	Z value	*p*-Value
Intercept across schools	295.33	65.74	4.49	<.0001
Slope across schools	43.94	9.70	4.53	<.0001
Intercept within schools	872.85	39.47	22.11	<.0001
Within student (residual)	374.66	10.29	36.40	<.0001

Goodness of fit

Criteria	Value
AIC (# of parameters)	-19385 (2)
BIC (# of parameters)	-19393 (2)

Note: Significant effects are shaded in gray.

Table A-9. HLM mathematics effects model with control variables only

Fixed effects

	Gamma coefficients	Standard error	*t*-Statistic	*p*-Value
Intercept	597.81	2.18	274.65	0.000
MFRLH980	-30.20	9.46	-3.19	0.002
CFRLH980	-13.40	2.36	-5.69	0.000
ENRL_970	0.00	0.02	0.07	0.942
Slope (RYEAR)	22.79	1.04	21.85	0.000
MFRLH980*RYEAR	4.32	4.68	0.92	0.355
CFRLH980*RYEAR	0.05	0.96	0.05	0.962
RYEAR*AVGENRL0	0.00	0.01	-0.37	0.712
RYEAR*MLONT970	-13.73	5.19	-2.65	0.008
RYEAR*LO_NAT97	-1.21	0.83	-1.46	0.145

Random effects

Parameter	Variance estimate	Standard error	Z value	*p*-Value
Intercept across schools	240.44	54.81	4.39	<.0001
Slope across schools	56.30	16.57	3.40	0.0003
Intercept within schools	823.01	40.56	20.29	<.0001
Within student (residual)	378.05	10.76	35.14	<.0001

Goodness of fit

Criteria	Value
AIC (# of parameters)	-19366 (10)
BIC (# of parameters)	-19382 (10)

Note: Significant effects are shaded in gray.

Table A-10. Level, year/grade, and direction of effect for teaching practices with significant effects when examined one at a time, 1 year at a time, for the best fit models

	Significant effect on third-grade score			Significant effect on learning rate		
Teaching practice	Student level[a]	School level I[b]	School level II[c]	Student level[a]	School level I[b]	School level II[c]
Exploration in instruction	Third-grade interaction with school poverty (++)[d]			Fifth grade (+++) & Fifth-grade interaction with bottom ¼ flag (—)		
Presentation and practice in instruction				Third grade (-)	Fifth grade (-) & Fifth-grade interaction with school poverty (–)	
Visibility of standards and assessments			1997 (++)	Fourth-grade interaction with school poverty (–)[d] & Fifth grade (++)	Fourth-grade interaction with school poverty (–)[d]	1997 (–)
Rating of professional development				Fifth grade (+++)	Fourth grade (+)	1998 (-)
Outreach to low achievers' parents						1998 interaction with bottom ¼ flag (++)[d]
Preparation for instruction			1997 (–)	Fifth grade (++)		Fifth grade (-)
Reported use of NCTM standards				Third-grade interaction with school poverty (+++)[d]		

[a] Student level represents the teaching practice of a student's individual teacher.

[b] School level I represents the average teaching practice of teachers of longitudinal students in a school.

[c] School level II represents the average teaching practice of all K–5 grade teachers in a school.

[d] Interaction was not in best fitting model but was significant at the 0.01 level in another model.

+	A positive significant effect was detected at the 0.05 level.
++	A positive significant effect was detected at the 0.01 level.
+++	A positive significant effect was detected at the 0.001 level.
-	A negative significant effect was detected at the 0.05 level.
–	A negative significant effect was detected at the 0.01 level.
—	A negative significant effect was detected at the 0.001 level.

BEST COPY AVAILABLE

Table A-11. HLM mathematics effects final model conditional on student and school characteristics and teaching practices

	Fixed effects			
	Gamma coefficients	Standard error	*t*-Statistic	*p*-Value
Intercept	597.12	2.16	276.38	0.000
MFRLH980	-27.86	9.25	-3.01	0.004
CFRLH980	-12.49	2.42	-5.15	0.000
ENRL_970	0.00	0.02	0.05	0.957
CINE1970	-0.01	0.01	-0.69	0.490
MINE1970	-0.03	0.02	-1.86	0.067
MFRLH980*CINE1970	0.09	0.03	2.72	0.007
MFRLH980*MINE1970	0.04	0.07	0.56	0.580
MSG_970	3.69	1.38	2.68	0.010
Slope (RYEAR)	23.12	0.79	29.40	0.000
MFRLH980*RYEAR	9.04	3.93	2.30	0.021
CFRLH980*RYEAR	-0.52	0.98	-0.53	0.595
RYEAR*MLONT970	-17.02	5.54	-3.07	0.002
RYEAR*LO_NAT97	-0.50	0.86	-0.58	0.559
RYEAR*AVGENRL0	0.00	0.01	-0.34	0.735
RYEAR*CINE1990	0.02	0.01	3.77	0.000
RYEAR*MINE1990	0.00	0.01	0.29	0.776
RYEAR*LO_NAT*CINE199	-0.03	0.01	-3.14	0.002
RYEAR*LO_NAT*MINE199	0.00	0.01	0.28	0.780
RYEAR*CINE2970	-0.01	0.00	-1.59	0.112
RYEAR*MINE2970	0.02	0.01	1.91	0.057
RYEAR*MINE2990	-0.01	0.01	-1.82	0.070
MFRLH9*RYEAR*MINE299	-0.12	0.03	-4.15	0.000
MSG_970*RYEAR	-0.18	0.80	-0.23	0.820
RYEAR*MING_980	-0.91	0.27	-3.38	0.001
RYEAR*CING_980	-0.17	0.15	-1.14	0.254
MFRLH9*RYEAR*CING_98	-1.45	0.49	-2.93	0.003
MFRLH9*RYEAR*MING_98	0.21	1.20	0.17	0.862
RYEAR*CING_990	0.30	0.17	1.81	0.071
RYEAR*MING_990	-0.45	0.30	-1.48	0.139
RYEAR*MIH2B980	0.17	0.08	2.23	0.026
RYEAR*MSH2B980	-0.19	0.28	-0.69	0.489
RYEAR*CIH2B990	0.24	0.07	3.50	0.001
RYEAR*MIH2B990	0.01	0.07	0.15	0.881
RYEAR*MSI_980	-1.38	0.82	-1.69	0.091
RYEAR*LO_NAT*MSI_980	3.38	0.88	3.84	0.000
RYEAR*MSL_970	0.29	0.70	0.42	0.674
RYEAR*CINL_990	0.10	0.19	0.52	0.600

Table A-11. HLM mathematics effects final model conditional on student and school characteristics and teaching practices (continued)

Fixed effects

	Gamma coefficients	Standard error	*t*-Statistic	*p*-Value
RYEAR*MINL_990	0.14	0.30	0.48	0.634
RYEAR*MSL_990	-1.32	0.60	-2.21	0.027
RYEAR*CQ31E970	-1.27	0.81	-1.58	0.115
RYEAR*MQ31E970	-0.02	1.63	-0.01	0.988
MFRLH9*RYEAR*CQ31E97	12.31	3.34	3.68	0.000
MFRLH9*RYEAR*MQ31E97	-30.56	8.32	-3.68	0.000

Random effects

Parameter	Variance estimate	Standard error	Z value	*p*-Value
Intercept across schools	218.18	51.06	4.27	<.0001
Slope across schools	20.78	8.54	2.43	0.0075
Intercept within schools	823.91	41.80	19.71	<.0001
Within student (residual)	368.18	10.83	34.01	<.0001

Goodness of fit

Criteria	Value
AIC (# of parameters)	-18121 (44)
BIC (# of parameters)	-18175 (44)

Note: Significant effects are shaded in gray.

Table A-12. Names and descriptions of variables used to predict each longitudinal student's third-grade score in the final HLM mathematics model

	Variable name
Control variables	
■ Student poverty	CFRLH980
■ School poverty	MFRLH980
■ School size in 1997	ENRL_970
Student-level variables	
■ Exploration in instruction	
- Third-grade teacher (1997)	CINE1970
- Third-grade teacher (1997), interacting with school poverty	MFRLH980*CINE1970
School-level variables	
■ Visibility of standards and assessments	
- Whole school (1997)	MSG_970
■ Exploration in instruction	
- All third-grade teachers (1997)	MINE1970
- All third-grade teachers (1997), interacting with school poverty	MFRLH980*MINE1970

Table A-13. Names and descriptions of variables used to predict each longitudinal student's learning rate in the final HLM mathematics model, control, and student-level variables

	Variable name
Control variables	
■ Student poverty	CFRLH980*RYEAR
■ School poverty	MFRLH980*RYEAR
■ Average school enrollment	RYEAR*AVGENRLO
■ Student's initial achievement status	RYEAR*LO_NAT97
■ School's initial achievement	RYEAR*MLONT970
Student-level variables	
■ Visibility of standards and assessments	
- Fourth-grade teacher (1998)	RYEAR*CING_980
- Fourth-grade teacher (1998), interacting with school poverty	MFRLH980*RYEAR*CING980
- Fifth-grade teacher (1999)	RYEAR*CING_990
■ Reported use of NCTM standards	
- Third-grade teacher (1997)	RYEAR*CQ31E970
- Third-grade teacher (1997), interacting with school poverty	MFRLH980*RYEAR*CQ31E970
■ Exploration in instruction	
- Fifth-grade teacher (1999)	RYEAR*CINE1990
- Fifth-grade teacher (1999), interacting with student's initial achievement status	RYEAR*LO_NAT*CINE1990
■ Presentation and practice in instruction	
- Third-grade teacher (1997)	RYEAR*CINE2970
■ Preparation for instruction	
- Fifth-grade teacher (1999)	RYEAR*CINL_990
■ Rating of professional development	
- Fifth-grade teacher (1999)	RYEAR*CIH2B990

Table A-13. Names and descriptions of variables used to predict each longitudinal student's learning rate in the final HLM mathematics model, control, and student-level variables (continued)

	Variable name
School-level variables	
■ Visibility of standards and assessments	
- All fourth-grade teachers (1998)	RYEAR*MING_980
- All fourth-grade teachers (1998), interacting with school poverty	MFRLH980* RYEAR*MING_980
- All fifth-grade teachers (1999)	RYEAR*MING_990
- Whole school (1997)	MSG_970*RYEAR
■ Reported use of NCTM standards	
- All third-grade teachers (1997)	RYEAR*MQ3IE970
- All third-grade teachers (1997), interacting with school poverty	MFRLH980* RYEAR*MQ3IE970
■ Exploration in instruction	
- All fifth-grade teachers (1999)	RYEAR*MINE1990
- All fifth-grade teachers (1999), interacting with student's initial achievement status	RYEAR*LO_NAT*MINE1990
■ Presentation and practice in instruction	
- All third-grade teachers (1997)	RYEAR*MINE2970
- All fifth-grade teachers (1999)	RYEAR*MINE2990
- All fifth-grade teachers (1999), interacting with school poverty	MFRHL980*RYEAR*MINE2990
■ Preparation for instruction	
- All fifth-grade teachers (1999)	RYEAR*MINL_990
- Whole school (1997)	RYEAR*MSL_970
- Whole school (1999)	RYEAR*MSL_990
■ Rating of professional development	
- All fourth-grade teachers (1998)	RYEAR*MIH2B980
- All fifth-grade teachers (1999)	RYEAR*MIH2B990
- Whole school (1998)	RYEAR*MSH2B980
■ Outreach to low achievers' parents	
- Whole school (1998)	RYEAR*MSI_980
- Whole school (1998), interacting with student's initial achievement status	RYEAR*LO_NAT*MSI_980

APPENDIX B

CHANGES IN THE TEACHER SURVEY ITEMS COMPRISING THE MEASURES OF STANDARDS-BASED REFORMS

Table B-1. Teacher responses to survey items that comprise the measures for reading

Teacher questionnaire item	Scale range	1997 mean	1999 mean
Visibility of standards and assessments	8–32	26.8	27.8[a]
Familiarity with:			
Content standards	1–4	3.24	3.45[a]
Curriculum frameworks	1–4	3.21	3.39[a]
Student assessments	1–4	3.45	3.58[a]
Performance standards	1–4	3.26	3.43[a]
Teaching reflects:			
Content standards	1–4	3.38	3.51[a]
Curriculum frameworks	1–4	3.36	3.44[a]
Student assessments	1–4	3.39	3.46[a]
Performance standards	1–4	3.32	3.42[a]
Basic instruction in the upper grades	0–540	171.4	182.6
Amount of time students:			
Read aloud	0–135	63.2	66.6
Complete reading workbooks or skill sheets	0–135	37.7	36.8
Practice phonics	0–135	32.0	35.3
Practice word attack	0–135	40.5	42.7
Preparation for instruction	5–20	17.9	18.0
How well teacher is prepared to:			
Use small group instruction	1–4	3.67	3.68
Take into account students' existing skill levels	1–4	3.62	3.63
Integrate reading into other content areas	1–4	3.62	3.62
Use a variety of assessment strategies	1–4	3.45	3.49
Teach heterogeneous groups	1–4	3.54	3.55
Rating of professional development [b]	0–41	23.9	25.8[a]
Quality of professional development in reading content	0–3	1.79	1.95[a]
Quality of professional development in instructional strategies	0–3	1.84	1.98[a]
Extent to which professional development activity was:			
Well matched to plans to change practice	0–5	3.12	3.30[a]
Designed to support reform efforts underway	0–5	3.15	3.35[a]
Designed to support standards or curriculum frameworks	0–5	3.26	3.45[a]
Designed to support state or district assessments	0–5	3.24	3.42[a]
Enhanced knowledge and skills in the following ways:			
Adapt teaching to meet state assessment requirements	0–5	2.99	3.12
Adapt teaching to meet standards/current framework requirements	0–5	3.00	3.14[a]
Gained confidence in new pedagogical approaches	0–5	2.75	2.90[a]

Table B-1. Teacher responses to survey items that comprise the measures for reading (continued)

Teacher questionnaire item	Scale range	1997 mean	1999 mean
Outreach to low achievers' parents	4–16	11.5	11.5
For how many of your low-achieving students did you:			
Initiate face-to-face meetings with parents	1–4	3.27	3.19
Initiate telephone calls when the child was having a problem	1–4	3.09	3.07
Initiate telephone calls when the child was not having a problem	1–4	2.32	2.37
Send materials home on ways parents can help their child	1–4	2.88	2.87

[a] Statistically significant difference between 1999 and 1997 at the 0.05 level

[b] For professional development items, comparison is 1998 to 1999 because many of these items were not asked in 1997

Table B-2. Teacher responses to survey items that comprise the measures for mathematics

Teacher questionnaire item	Scale range	1997 mean	1999 mean
Visibility of standards and assessments	8–32	26.8	27.7[a]
Familiarity with:			
Content standards	1–4	3.31	3.46[a]
Curriculum frameworks	1–4	3.26	3.39[a]
Student assessments	1–4	3.40	3.53[a]
Performance standards	1–4	3:26	3.42[a]
Teaching reflects:			
Content standards	1–4	3.41	3.49[a]
Curriculum frameworks	1–4	3.36	3.43[a]
Student assessments	1–4	3.35	3.44[a]
Performance standards	1–4	3.28	3.39[a]
Exploration in instruction	0–1,215	274.6	286.9
Amount of time teachers:			
Use manipulatives to demonstrate a concept	0–135	57.51	57.97
Discuss multiple approaches to solving a problem	0–135	49.61	50.79
Amount of time students:			
Work on problems in small groups to find a joint solution	0–135	34.36	35.05
Have whole class discuss solutions developed in small groups	0–135	26.91	25.87
Have student-led whole group discussions	0–135	13.34	15.86[a]
Represent and analyze relationships using tables and graphs	0–135	17.42	17.42
Respond to questions or assignments that required writing at least a paragraph	0–135	13.44	16.33[a]
Work with manipulatives	0–135	56.75	54.76
Work on projects/assignments that take a week or more to finish	0–135	6.70	9.32[a]
Presentation and practice	0–945	330.8	362.7[a]
Amount of time teachers:			
Lecture or present	0–135	34.40	52.16[a]
Lead whole group discussions	0–135	59.32	59.80
Demonstrate working an exercise at a board	0–135	57.42	56.64
Administer a test	0–135	16.15	16.09
Amount of time students:			
Respond orally to questions on subject matter	0–135	64.76	73.65[a]
Work individually on written assignments/worksheets in class	0–135	56.37	59.33
Practice or drill on computational skills	0–135	42.48	41.55

BEST COPY AVAILABLE

Table B-2. Teacher responses to survey items that comprise the measures for mathematics (continued)

Teacher questionnaire item	Scale range	1997 mean	1999 mean
Preparation for instruction	8–32	27.6	27.8
How well teacher is prepared to:			
Present applications of mathematics concepts	1–4	3.59	3.60
Use cooperative learning groups in mathematics	1–4	3.40	3.43
Take into account students' previous conceptions about mathematics when planning curriculum and instruction	1–4	3.42	3.42
Integrate mathematics with other subject areas	1–4	3.33	3.36
Manage a class of students who are using manipulatives	1–4	3.57	3.60
Use a variety of assessment strategies	1–4	3.35	3.42[a]
Use the textbook as a resource rather than as the primary instruction tool	1–4	3.42	3.45
Teach groups that are heterogeneous in ability	1–4	3.48	3.51
Rating of professional development[b]	0–41	23.0	21.6
Quality of professional development in mathematics content	0–3	1.74	1.62[a]
Quality of professional development in instructional strategies	0–3	1.76	1.63[a]
Extent to which professional development activity was:			
Well matched to plans to change practice	0–5	3.03	2.89
Designed to support reform efforts underway	0–5	3.07	2.94
Designed to support standards or curriculum frameworks	0–5	3.19	3.05
Designed to support state or district assessments	0–5	3.17	3.00[a]
Enhanced knowledge and skills in the following ways:			
Adapt teaching to meet state assessment requirements	0–5	2.94	2.76[a]
Adapt teaching to meet standards/current framework requirements	0–5	2.95	2.77[a]
Gained confidence in new pedagogical approaches	0–5	2.67	2.39[a]
Outreach to low achievers' parents	4–16	11.5	11.5
For how many of your low-achieving students did you:			
Initiate face-to-face meetings with parents	1–4	3.28	3.22
Initiate telephone calls when the child was having a problem	1–4	3.11	3.09
Initiate telephone calls when the child was not having a problem	1–4	2.30	2.34
Send materials home on ways parents can help their child	1–4	2.88	2.90

[a] Statistically significant difference between 1999 and 1997 at the 0.05 level

[b] For professional development items, comparison is 1998 to 1999 because many of these items were not asked in 1997

APPENDIX C

ADDITIONAL ANALYSES OF CHANGES

Analysis of Reading Scores of Consecutive Fourth Grades

The hierarchical linear modeling (HLM) analysis focused on the achievement of the longitudinal students, who were tested in the third grade in 1997, the fourth grade in 1998, and in the fifth grade in 1999. We also collected student test data from three consecutive fourth grades (1997 through 1999) in the study schools. These data are a rich set of information about fourth-grade achievement for almost 9,000 students. We analyzed these data to address two questions:

- What are the characteristics of students that statistically relate to test scores?
- What instructional indices, if any, show a relationship to test scores when these individual student characteristics are held constant?

For the first question, we pooled data from all 3 years on fourth graders for whom we had test scores. For each student who took the closed-ended reading test in 1997, 1998, or 1999, we gathered information on the following items:

- SAT-9 test score;
- Year the student took the test (1997, 1998, or 1999);
- Student's sex;
- Student's race/ethnicity;
- Whether the student was eligible for free or reduced-priced lunch;
- Whether the student had an Individual Education Plan (IEP);
- Whether the school classified the student as limited-English proficient (LEP); and
- Whether the student was in a state that had a high-stakes, fourth-grade assessment in reading in the year they took the test.

We fitted an ordinary least squares (OLS) regression model, where the dependent variable was test score and the independent variables consisted of those listed above. The results of these regressions are summarized in Table C-1, where coefficients listed were significant at the 0.01 level or better.

Table C-1. Results of reading regressions using consecutive fourth grades[1]

Variable	Closed-ended reading (N=8,952)
Intercept	644.9
1998 versus 1997	(ns)
1999 versus 1997	(ns)
Male versus female	-4.9
African American versus white	-19.1
Hispanic versus white	-8.2
Asian versus white	(ns)
Student has an IEP	-26.2
Student classified as LEP	-16.9
Student eligible for free or reduced-price lunch	-16.4
Student in high-stakes state or district	8.2

[1] Cell entries are the coefficients from the regressions; (ns) means the coefficient was not significantly different than zero at the 0.01 level or better.

The significant effects, then, were the following:

- Males scored lower than females;
- African American students scored lower than white students;
- Hispanic students scored lower than white, non-Hispanic students;
- Students with IEPs scored lower than others;
- Students classified as LEP scored lower than others;
- Students eligible for free or reduced-price lunchs scored lower than others; and
- Students living in a state or district where a high-stakes assessment is administered in the fourth grade scored higher than others.

For each student and for each year, we calculated the difference between what the student actually scored on the test and what the equation predicted he or she would score. If this difference was a positive number, then the student scored better than predicted; if there was a negative difference, the student scored worse than predicted.

From these student-level differences, we created a measure of each Longitudinal Evaluation of School Change and Performance (LESCP) school's performance, by year, by averaging the difference between predicted and actual fourth-grade scores across all the test takers in the school. Compared with just using the mean school score, this measure allows us to compare schools while controlling for the differences in student characteristics shown to be related to test scores.

We next sought to relate this measure of school performance with all the school-level (K–5) measures that were used in the HLM analyses (see Boxes 3–1 through 3–5 in Chapter 3). We used linear regression to correlate the school-level indices with the performance measures described above for each year of the study. The results for the closed ended reading test are summarized in Table C-2.

Table C-2. Relationships between school-level indices and school performance

School-level index	1997 Reading	1998 Reading	1999 Reading
Visibility of standards and assessments			+
Basic instruction in upper grades			
Rating of professional development			
Outreach to low achievers' parents			
Preparation for instruction			

Note: + indicates a positive relationship significant at the .05 level or better

The only significant relationship uncovered was a positive one, between the visibility of standards and assessments and 1999 school performance. This index also showed positive effects on third-grade performance in the HLM analysis, lending some support to that finding.

Analysis of Schools' Gains on State Tests

We also used teacher indices to predict school-level results on state assessments. Our outcome measure for this analysis was a very simple one, providing a number that could be compared in at least a rough way across schools in different states. It was the ratio of 1999 test score to 1997 test score on the state assessment for the same grade within a school—for example, the ratio of the average test score of fifth graders in 1999 to the average test score of fifth graders in 1997. Six LESCP states assessed their students at the same grade level in 1997 and in 1999 in reading and/or mathematics. For the schools in these states, we computed the average ratio of 1999 test score to 1997 test score for any and all grade levels tested, so that each state had one score per subject. For example, one state administered its reading assessments at both the third- and fifth-grade levels. We computed the ratio of 1999 to 1997 test scores for third and fifth grade separately and then took the average of the ratios to come up with that school's reading ratio.

We predicted the school's ratio with school poverty and with the average value for each index of instruction or instructional support in the school based on all K–5 teachers (described in Boxes 3–1 through 3–5 in Chapter 3). Groupings of these measures 1 year at a time gave us separate analyses for 1997, 1998, and 1999. Thus, we predicted the school's ratio for state assessment in reading using 1997, 1998, and 1999 measures of the following:

- School poverty;
- Visibility of standards and assessments;
- Basic instruction in upper grades;
- Preparation for instruction;
- Rating of professional development; and
- Outreach to low achievers' parents.

In contrast to our HLM findings, however, this analysis did not yield any significant relationships between the variables listed above and the reading state assessment ratio.

Analysis of Mathematics Scores of Consecutive Fourth Grades

As with our reading analyses, we supplemented our HLM analysis of the mathematics achievement of the longitudinal students who were tested in third, fourth, and fifth grades (1997 through 1999) with student test data from three consecutive fourth grades (1997 through 1999) in the study schools. These fourth grade achievement data enabled us to explore two questions:

- What are the characteristics of students that statistically relate to mathematics test scores?
- What instructional indices, if any, show a relationship to test scores when these individual student characteristics are held constant?

In a parallel process to our reading analysis, we first formed a record in a data set for each student who took the closed-ended mathematics test in 1997, 1998, or 1999. The information included for each student was the following:

- SAT-9 test score;
- Year the student took the test (1997, 1998, or 1999);
- Student's sex;
- Student's race/ethnicity;
- Whether the student was eligible for free or reduced-price lunch;
- Whether the student had an IEP;
- Whether the school classified the student as LEP; and
- Whether the student was in a state that had a high-stakes, fourth-grade assessment in mathematics in the year they took the test.

We then fit an OLS regression model, using the test score as the dependent variable and the variables listed above as the independent variables. Table C-3 summarizes the results of these regressions, displaying coefficients that were significant at the 0.01 level or better.

As indicated in Table C-3, the significant effects for mathematics scores were the following:

- African American students scored lower than white students;
- Asian students scored higher than white students;
- Students with IEPs scored lower than others;
- Students classified as LEP scored lower than others;
- Students eligible for free or reduced-price lunch scored lower than others; and
- Students living in a state or district where a high stakes assessment is administered in the fourth grade scored higher than others.

Table C-3. Results of mathematics regressions using consecutive fourth grades[1]

Variable	Closed-ended mathematics (N=8,688)
Intercept	627.1
1998 versus 1997	(ns)
1999 versus 1997	(ns)
Male versus female	(ns)
African American versus white	-13.4
Hispanic versus white	(ns)
Asian versus white	12.8
Student has an IEP	-28.1
Student classified as LEP	-8.5
Student eligible for free or reduced-price lunch	-12.9
Student in high-stakes state or district	13.6

[1] Cell entries are the coefficients from the regressions; (ns) means the coefficient was not significantly different from zero at the 0.01 level or better.

We calculated the difference between what a student actually scored on the test and the score predicted by the regression equation. The student scored better than predicted by the equation if this difference was positive; if the difference was a negative number, the student scored worse than predicted.

For each year, we used this difference to create a measure of each LESCP school's performance on the closed-ended mathematics test by averaging these individual student differences between actual and predicted fourth grade scores across all test takers in the school. This method allowed us to compare schools while taking into account the differences in student characteristics that are significantly related to mathematics test scores.

Our next step was to relate this measure of school performance with school-level measures based on all K–5 teachers of all the instructional indices that were included in the HLM analysis previously described in Chapter 3. Using linear regression, we correlated the school-level indices with the performance measures described above for each year of the study. The relationships found for the closed-ended mathematics test are displayed in Table C-4.

Table C-4. Relationships between school-level indices and school performance

School-level index	1997 Mathematics	1998 Mathematics	1999 Mathematics
Visibility of standards and assessments	+	+	+
Exploration in instruction	-		
Presentation and practice in instruction			
Preparation for instruction			
Rating of professional development			
Outreach to low achievers' parents	-		

Note: + indicates a positive relationship significant at the .05 level or better
- indicates a negative relationship significant at the .05 level or better

This regression analysis showed a significant positive relationship between the visibility of standards and assessments and school performance in each year of the study (1997 through 1999). This is consistent with the finding in the HLM analysis that this index when measured in 1997 had a positive relationship to third-grade mathematics scores.

The regression also found negative relationships between 1997 scores and two additional variables that did not show main effects on third-grade scores in the HLM model: exploration in instruction and outreach to low achievers' parents.

Analysis of Schools' Gains on State Tests

As in the reading analysis, we also used teacher indices to predict school-level results on state assessments in mathematics. Again, for the six out of seven LESCP states that assessed their students at the same grade level in 1997 and in 1999 in mathematics, we used the ratio of 1999 test score to 1997 test score on the state assessments for the same grade within a school as our outcome measure. For schools in these states, the average ratio was computed for any and all grade levels tested, resulting in just one score for each school in mathematics.

We predicted each school's ratio using school poverty and the average value for each teacher index included in the HLM analyses (described in Boxes 3–7 to 3–12 in Chapter 3). This created three separate analyses for the 1997 through 1999 values on these indices. The school's ratio for state assessment in mathematics was therefore predicted using 1997, 1998, and 1999 measures of the following items:

- School poverty;
- Visibility of standards and assessments;
- Exploration in instruction;
- Presentation and practice in instruction;
- Preparation for instruction;
- Rating of professional development; and
- Outreach to low achievers' parents.

Only the regression for the 1997 measures of the above variables against the mathematics state assessment ratio yielded a significant relationship. Together, the 1997 measures for these variables explained 19.7 percent of the variation in state assessment ratios.

The 1997 measure of exploration in instruction for all teachers in grades K–5 had a significant positive association with the state assessment ratio, at a *p*-value of 0.030. This suggests that schools where teachers rated high on our exploration in instruction index in 1997, on average, demonstrated a statistically significant increase in scores on the state assessment in mathematics between 1997 and 1999. This analysis complements the HLM finding that exploration in instruction by a student's fifth-grade teacher had a positive significant relationship with that student's gain in score from 1997-1999, as shown in Table C-5 below. Although the variable proved significant at different levels in the two analyses, the direction and significance of the relationship are consistent.

Table C-5. Comparison of findings for changes in mathematics scores for HLM and regression analysis using 1997 exploration in instruction measure

Exploration in instruction index level	HLM finding re. 1997-1999 change in mathematics score on SAT-9	Regression finding re. 1997/1999 ratio for mathematics state assessment
All teachers in the school (1997)	ns	+
Student's fifth-grade teacher (1999)	+	NA

Note: + indicates a positive relationship significant at the 0.05 level or better
NA indicates that the level for the index is not applicable for that test
ns indicates the difference was not statistically different

APPENDIX D

RESULTS OF RELIABILITY ANALYSES FOR INDEX CONSTRUCTION

Table D-1. Visibility of standards and assessments (reading)

Item no.	Teacher survey items	Item correlation with index		
		1997	1998	1999
5a	Please indicate how familiar you are with **content standards in reading/language arts** for your state or district. (scale: 1–4)	.71	.71	.67
5b	Please indicate how familiar you are with **curriculum frameworks in reading/language arts** for your state or district. (scale: 1–4)	.69	.68	.65
5c	Please indicate how familiar you are with **student assessments in reading/language arts** for your state or district. (scale: 1–4)	.66	.64	.60
5d	Please indicate how familiar you are with **performance standards in reading/language arts** for your state or district. (scale: 1–4)	.69	.69	.67
6a	To what extent do you believe the **reading/language arts curriculum** you teach reflects **content standards** in reading/language arts? (scale: 1–5)	.66	.71	.67
6b	To what extent do you believe the **reading/language arts curriculum** you teach reflects **curriculum frameworks** in reading/language arts? (scale: 1–5)	.66	.71	.68
6c	To what extent do you believe the **reading/language arts curriculum** you teach reflects **student assessments** in reading/language arts? (scale: 1–5)	.66	.68	.66
6d	To what extent do you believe the **reading/language arts curriculum** you teach reflects **performance standards** in reading/language arts? (scale: 1–5)	.66	.68	.68
Cronbach's alpha coefficient for index		.89	.90	.89

Table D-2. Basic instruction in the upper grades (reading)

Item no.	Teacher survey items	Item correlation with index		
		1997	1998	1999
13d	Indicate how frequently (and how much per lesson) you have your students **read aloud.** (scale: 0–135)	.36	.38	.42
13g	Indicate how frequently (and how much per lesson) you have your students **complete reading workbooks or skill-sheet assignments.** (scale: 0–135)	.41	.44	.45
13h	Indicate how frequently (and how much per lesson) you have your students **practice phonics.** (scale: 0–135)	.62	.67	.69
13i	Indicate how frequently (and how much per lesson) you have your students **practice word attack.** (scale: 0–135)	.63	.68	.72
Cronbach's alpha coefficient for index		.71	.74	.76

BEST COPY AVAILABLE

Table D-3. Preparation for instruction (reading)

Item no.	Teacher survey items	Item correlation with index		
		1997	1998	1999
7a	How well are you prepared to **use small group instruction for reading/language arts?** (scale: 1–4)	.60	.65	.68
7b	How well are you prepared to **take into account students' existing skills levels when planning curriculum and instruction?** (scale: 1–4)	.68	.71	.74
7c	How well are you prepared to **integrate reading/language arts into other content areas?** (scale: 1–4)	.61	.65	.66
7d	How well are you prepared to **use a variety of assessment strategies?** (scale: 1–4)	.66	.67	.69
7e	How well are you prepared to **teach groups that are heterogeneous in ability?** (scale: 1–4)	.66	.66	.68
Cronbach's alpha coefficient for index		.84	.85	.87

Table D-4. Rating of professional development (reading)

	Teacher survey items	Item correlation with index	
		1998	1999
57aqual	If you have received any professional development in the past 12 months on **content in reading**, what was the quality? (scale: 1–3)	.60	.52
57cqual	If you have received any professional development in the past 12 months on **instructional strategies for teaching reading**, what was the quality? (scale: 1–3)	.59	.52
58aa	To what extent was the professional development activity **well matched to your school's or department's plan to change practice?** (scale: 1–5)	.77	.75
58ab	To what extent was the professional development activity **designed to support reform efforts underway in your school?** (scale: 1–5)	.76	.78
58ac	To what extent was the professional development activity **designed to support state or district standards or curriculum frameworks?** (scale: 1–5)	.81	.76
58ad	To what extent was the professional development activity **designed to support state or district assessment?** (scale: 1–5)	.79	.74
58ba	To what extent do you feel that your knowledge and skills have been enhanced by **helping you adapt your teaching to meet state assessment requirements** as a result of your participation in the professional development experiences you have had in the past year? (scale: 1–5)	.80	.73
58bb	To what extent do you feel that your knowledge and skills have been enhanced by **helping you adapt your teaching to meet state standards or curriculum framework requirements** as a result of your participation in the professional development experiences you have had in the past year? (scale: 1–5)	.80	.77
58be	To what extent do you feel that your knowledge and skills have been enhanced by **gaining confidence in using new pedagogical approaches in teaching reading/English language arts** as a result of your participation in the professional development experiences you have had in the past year? (scale: 1–5)	.71	.59
Cronbach's alpha coefficient for index		.93	.91

Note: This scale can only be constructed for 1998 and 1999. In 1997, we did not collect information on all the items contained in this scale.

BEST COPY AVAILABLE

Table D-5. Visibility of standards and assessments (mathematics)

Teacher survey items		Item correlation with index		
		1997	1998	1999
30a	Please indicate how familiar you are with **content standards in mathematics** for your state or district. (scale: 1–4)	.72	.71	.70
30b	Please indicate how familiar you are with **curriculum frameworks in mathematics** for your state or district. (scale: 1–4)	.69	.70	.73
30c	Please indicate how familiar you are with **student assessments in mathematics** for your state or district. (scale: 1–4)	.70	.66	.71
30d	Please indicate how familiar you are with **performance standards in mathematics** for your state or district. (scale: 1–4)	.72	.70	.73
31a	To what extent do you believe the **mathematics curriculum** you teach reflects content standards in mathematics? (scale: 1–5)	.69	.70	.73
31b	To what extent do you believe the **mathematics curriculum** you teach reflects curriculum frameworks in mathematics? (scale: 1–5)	.66	.71	.72
31c	To what extent do you believe the **mathematics curriculum** you teach reflects **student assessments** in mathematics? (scale: 1–5)	.69	.65	.71
31d	To what extent do you believe the **mathematics curriculum** you teach reflects **performance standards** in mathematics? (scale: 1–5)	.70	.68	.74
Cronbach's alpha coefficient for index		.90	.90	.91

Table D-6. Exploration in instruction (mathematics)

Item no.	Teacher survey items	Item correlation with index		
		1997	1998	1999
45e	Indicate the extent to which (and how much) you **use manipulatives to demonstrate a concept.** (scale: 0–135)	.50	.54	.52
45f	Indicate the extent to which (and how much) you **discuss multiple approaches to solving a problem.** (scale: 0–135)	.48	.50	.51
46c	Indicate how frequently (and how much per lesson) you have your students **work on problems in small group to find a joint solution.** (scale: 0–135)	.55	.47	.54
46d	Indicate how frequently (and how much per lesson) you have your students **have whole class discuss solutions developed in small groups.** (scale: 0–135)	.58	.59	.57
46f	Indicate how frequently (and how much per lesson) you have your students **have student-led whole group discussions.** (scale: 0–135)	.54	.47	.50
46g	Indicate how frequently (and how much per lesson) you have your students **represent and analyze relationships using tables and graphs.** (scale: 0–135)	.51	.35	.48
46i	Indicate how frequently (and how much per lesson) you have your students **respond to questions or assignments that required writing at least a paragraph.** (scale: 0–135)	.43	.26	.37
46j	Indicate how frequently (and how much per lesson) you have your students **work with manipulatives.** (scale: 0–135)	.50	.49	.47
46k	Indicate how frequently (and how much per lesson) you have your students **work on projects/assignments that take a week or more to finish.** (scale: 0–135)	.46	.26	.36
Cronbach's alpha coefficient for index		.80	.75	.78

BEST COPY AVAILABLE

Table D-7. Presentation and practice in instruction (mathematics)

Item no.	Teacher survey items	Item correlation with index		
		1997	1998	1999
45a	Indicate the extent to which (and how much) you **lecture or present.** (scale: 0–135)	.49	.45	.58
45b	Indicate the extent to which (and how much) **lead whole group discussions.** (scale: 0–135)	.39	.43	.57
45c	Indicate the extent to which (and how much) **demonstrate working an exercise at the board.** (scale: 0–135)	.51	.54	.60
45d	Indicate the extent to which (and how much) **administer a test.** (scale: 0–135)	.40	.35	.42
46a	Indicate how frequently (and how much per lesson) you have your students **respond orally to questions on subject matter.** (scale: 0–135)	.42	.36	.47
46b	Indicate how frequently (and how much per lesson) you have your students **work individually on written assignments or worksheets in class.** (scale: 0–135)	.42	.38	.54
46e	Indicate how frequently (and how much per lesson) you have your students **practice or drill on computational skills.** (scale: 0–135)	.47	.41	.49
Cronbach's alpha coefficient for index		.72	.70	.79

Table D-8. Preparation for instruction (mathematics)

Item no.	Teacher survey items	Item correlation with index		
		1997	1998	1999
32a	How well prepared are you to **present the applications of mathematics concepts?** (scale: 1–4)	.64	.68	.74
32b	How well prepared are you to **use cooperative learning groups in mathematics?** (scale: 1–4)	.67	.71	.71
32c	How well prepared are you to **take into account students' prior conceptions about mathematics when planning curriculum and instruction?** (scale: 1–4)	.67	.70	.78
32d	How well prepared are you to **integrate mathematics with other subject areas?** (scale: 1–4)	.68	.69	.71
32e	How well prepared are you to **manage a class of students who are using manipulatives?** (scale: 1–4)	.73	.71	.75
32f	How well prepared are you to **use a variety of assessment strategies?** (scale: 1–4)	.75	.74	.77
32g	How well prepared are you to **use the textbook as a resource rather than as the primary instructional tool?** (scale: 1–4)	.65	.63	.70
32h	How well prepared are you to **teach groups that are heterogeneous in ability?** (scale: 1–4)	.67	.73	.76
Cronbach's alpha coefficient for index		.90	.90	.92

Table D-9. Rating of professional development (mathematics)

Teacher survey items		Item correlation with index	
		1998	1999
57bqual	If you have received any professional development in the past 12 months on **content in mathematics**, what was the quality? (scale: 1–3)	.55	.47
57dqual	If you have received any professional development in the past 12 months on **instructional strategies for teaching mathematics**, what was the quality? (scale: 1–3)	.56	.46
58aa	To what extent was the professional development activity **well matched to your school's or department's plan to change practice?** (scale: 1–5)	.74	.73
58ab	To what extent was the professional development activity **designed to support reform efforts underway in your school?** (scale: 1–5)	.75	.75
58ac	To what extent was the professional development activity **designed to support state or district standards or curriculum frameworks?** (scale: 1–5)	.80	.77
58ad	To what extent was the professional development activity **designed to support state or district assessment?** (scale: 1–5)	.77	.73
58ba	To what extent do you feel that your knowledge and skills have been enhanced by **helping you adapt your teaching to meet state assessment requirements** as a result of your participation in the professional development experiences you have had in the past year? (scale: 1–5)	.80	.73
58bb	To what extent do you feel that your knowledge and skills have been enhanced by **helping you adapt your teaching to meet state standards or curriculum framework requirements** as a result of your participation in the professional development experiences you have had in the past year? (scale: 1–5)	.80	.78
58bd	To what extent do you feel that your knowledge and skills have been enhanced by **gaining confidence in using new pedagogical approaches in teaching mathematics** as a result of your participation in the professional development experiences you have had in the past year? (scale: 1–5)	.63	.58
Cronbach's alpha coefficient for index		.92	.90

Note: This scale can only be constructed for 1998 and 1999; in 1997 we did not collect information on all the items contained in this scale.

Table D-10. Outreach to low achievers' parents

Item no.	Teacher survey items	Item correlation with index		
		1997	1998	1999
60a	For how many parents of low-achieving students did you **initiate face-to-face meetings?** (scale: 1–4)	.49	.48	.51
60b	For how many parents of low-achieving students did **initiate telephone calls to parents when their child was having problems?** (scale: 1–4)	.63	.63	.64
60c	For how many parents of low-achieving students did **initiate telephone calls to parents when their child was not having problems?** (scale: 1–4)	.51	.55	.55
60d	For how many parents of low-achieving students did **send materials home to parents on ways they can help their child at home?** (scale: 1–4)	.47	.46	.48
Cronbach's alpha coefficient for index		.73	.73	.75

BEST COPY AVAILABLE